NICOLAS BOULEAU
DOMINIQUE LÉPINGLE

NUMERICAL METHODS FOR STOCHASTIC PROCESSES

BOULEAU LÉPINGLE
NUMERICAL METHODS FOR STOCHASTIC PROCESSES

519.2
B60

WILEY SERIES IN PROBABILITY AND MATHEMATICAL STATISTICS

NICOLAS BOULEAU
DOMINIQUE LÉPINGLE

NUMERICAL METHODS FOR STOCHASTIC PROCESSES

WILEY SERIES IN PROBABILITY AND MATHEMATICAL STATISTICS

WILEY SERIES IN PROBABILITY
AND MATHEMATICAL STATISTICS

ESTABLISHED BY WALTER A. SHEWHART AND SAMUEL S. WILKS
Editors
Vic Barnett, Ralph A. Bradley, Nicholas I. Fisher, J. Stuart Hunter, J. B. Kadane, David G. Kendall, Adrian F. M. Smith, Stephen M. Stigler, Jozef L. Teugels, Geoffrey S. Watson

Probability and Mathematical Statistics

ANDERSON • An Introduction to Multivariate Statistical Analysis, *Second Edition*
ARNOLD, BALAKRISHNAN, and NAGARAJA • A First Course in Order Statistics
BACCELLI, COHEN, OLSDER, and QUADRAT • Synchronization and Linearity: An Algebra for Discrete Event Systems
BARNETT • Comparative Statistical Inference, *Second Edition*
BERNARDO and SMITH • Bayesian Statistical Concepts and Theory
BHATTACHARYYA and JOHNSON • Statistical Concepts and Methods
BILLINGSLEY • Probability and Measure, *Second Edition*
BOROVKOV • Asymptotic Methods in Queuing Theory
BRANDT, FRANKEN, and LISEK • Stationary Stochastic Models
CAINES • Linear Stochastic Systems
CHEN • Recursive Estimation and Control for Stochastic Systems
CONSTANTINE • Combinatorial Theory and Statistical Design
COVER and THOMAS • Elements of Information Theory
*DOOB • Stochastic Processes
DUDEWICZ and MISHRA • Modern Mathematical Statistics
ETHIER and KURTZ • Markov Processes: Characterization and Convergence
FELLER • An Introduction to Probability Theory and Its Applications, Volume I, *Third Edition*, Revised; Volume II, *Second Edition*
FULLER • Introduction to Statistical Time Series
FULLER • Measurement Error Models
GIFI • Nonlinear Multivariate Analysis
GUTTORP • Statistical Inference for Branching Processes
HALD • A History of Probability and Statistics and Their Applications before 1750
HALL • Introduction to the Theory of Coverage Processes
HANNAN and DEISTLER • The Statistical Theory of Linear Systems
HEDAYAT and SINHA • Design and Inference in Finite Population Sampling
HOEL • Introduction to Mathematical Statistics, *Fifth Edition*
HUBER • Robust Statistics
IMAN and CONOVER • A Modern Approach to Statistics
KAUFMAN and ROUSSEEUW • Finding Groups in Data: An Introduction to Cluster Analysis
LARSON • Introduction to Probability Theory and Statistical Inference, *Third Edition*
LESSLER and KALSBEEK • Nonsampling Error in Surveys
MORGENTHALER and TUKEY • Configural Polysampling: A Route to Practical Robustness
MUIRHEAD • Aspects of Multivariate Statistical Theory
OLIVER and SMITH • Influence Diagrams, Belief Nets and Decision Analysis
*PARZEN • Modern Probability Theory and Its Applications
PILZ • Bayesian Estimation and Experimental Design in Linear Regression Models
PRESS • Bayesian Statistics: Principles, Models, and Applications
PUKELSHEIM • Optimal Experimental Design
PURI and SEN • Nonparametric Methods in General Linear Models
PURI, VILAPLANA, and WERTZ • New Perspectives in Theoretical and Applied Statistics

*Now available in a lower priced paperback edition in the Wiley Classics Library.

Probability and Mathematical Statistics (Continued)

RAO • Asymptotic Theory of Statistical Inference
RAO • Linear Statistical Inference and Its Applications, *Second Edition*
ROBERTSON, WRIGHT, and DYKSTRA • Order Restricted Statistical Inference
ROGERS and WILLIAMS • Diffusions, Markov Processes, and Martingales, Volume II: Îto Calculus
ROHATGI • A Introduction to Probability Theory and Mathematical Statistics
ROSS • Stochastic Processes
RUBINSTEIN • Simulation and the Monte Carlo Method
RUZSA and SZEKELY • Algebraic Probability Theory
SCHEFFE • The Analysis of Variance
SEBER • Linear Regression Analysis
SEBER • Multivariate Observations
SEBER and WILD • Nonlinear Regression
SERFLING • Approximation Theorems of Mathematical Statistics
SHORACK and WELLNER • Empirical Processes with Applications to Statistics
STAUDTE and SHEATHER • Robust Estimation and Testing
STOYANOV • Counterexamples in Probability
STYAN • The Collected Papers of T. W. Anderson: 1943–1985
WHITTAKER • Graphical Models in Applied Multivariate Statistics
YANG • The Construction Theory of Denumerable Markov Processes

Applied Probability and Statistics

ABRAHAM and LEDOLTER • Statistical Methods for Forecasting
AGRESTI • Analysis of Ordinal Categorical Data
AGRESTI • Categorical Data Analysis
ANDERSON and LOYNES • The Teaching of Practical Statistics
ANDERSON, AUQUIER, HAUCK, OAKES, VANDAELE, and WEISBERG • Statistical Methods for Comparative Studies
ASMUSSEN • Applied Probability and Queues
*BAILEY • The Elements of Stochastic Processes with Applications to the Natural Sciences
BARNETT • Interpreting Multivariate Data
BARNETT and LEWIS • Outliers in Statistical Data, *Second Edition*
BARTHOLOMEW, FORBES, and McLEAN • Statistical Techniques for Manpower Planning, *Second Edition*
BATES and WATTS • Nonlinear Regression Analysis and Its Applications
BELSLEY • Conditioning Diagnostics: Collinearity and Weak Data in Regression
BELSLEY, KUH, and WELSCH • Regression Diagnostics: Identifying Influential Data and Sources of Collinearity
BHAT • Elements of Applied Stochastic Processes, *Second Edition*
BHATTACHARYA and WAYMIRE • Stochastic Processes with Applications
BIEMER, GROVES, LYBERG, MATHIOWETZ, and SUDMAN • Measurement Errors in Surveys
BIRKES and DODGE • Alternative Methods of Regression
BLOOMFIELD • Fourier Analysis of Time Series: An Introduction
BOLLEN • Structural Equations with Latent Variables
BOULEAU • Numerical Methods for Stochastic Processes
BOX • R. A. Fisher, the Life of a Scientist
BOX and DRAPER • Empirical Model-Building and Response Surfaces
BOX and DRAPER • Evolutionary Operation: A Statistical Method for Process Improvement
BOX, HUNTER, and HUNTER • Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building
BROWN and HOLLANDER • Statistics: A Biomedical Introduction
BUCKLEW • Large Deviation Techniques in Decision, Simulation, and Estimation

Continued on back end papers

*Now available in a lower priced paperback edition in the Wiley Classics Library.



Numerical Methods for Stochastic Processes

LIBRARY

Please return this book or have the loan renewed on or before the date stamped below

GLASGOW
CALEDONIAN
UNIVERSITY

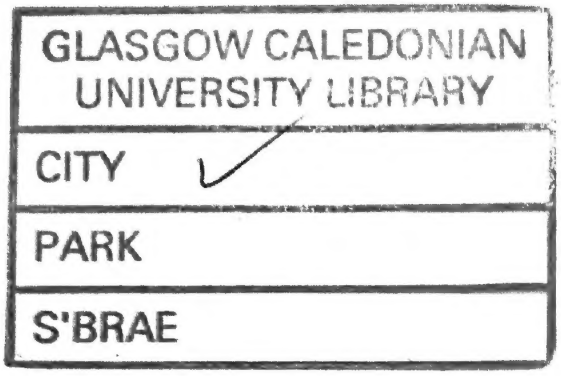
Numerical Methods for Stochastic Processes

NICOLAS BOULEAU
Ecole Nationale des Ponts et Chaussées

DOMINIQUE LÉPINGLE
Université d'Orléans



A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York • Chichester • Brisbane • Toronto • Singapore


GLASGOW CALEDONIAN
UNIVERSITY LIBRARY
CITY
PARK
S'BRAE

519.2
B60

This text is printed on acid-free paper.

Copyright © 1994 by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work beyond that permitted by Section 107 or 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012.

Library of Congress Cataloging in Publication Data:

Bouleau, Nicolas.

Numerical methods for stochastic process / by Nicolas Bouleau and Dominique Lépingle.

p. cm. — (Wiley series in probability and mathematical statistics. Applied probability and statistics)

"A Wiley-Interscience publication."

Includes bibliographical references and index.

ISBN 0-471-54641-0 (acid-free)

1. Stochastic processes–Mathematical models. 2. Monte Carlo method. I. Lépingle, Dominique. II. Title. III. Series.

QA274.B67 1993

519.2–dc20

93-10302

Printed in the United States of America

10 9 8 7 6 5 4 3

Dedicated to Jacques Neveu for his 60th birthday

Contents

Preface xiii

1. Preliminaries 1

A Set Theory and General Topology 1
A.1 Some Notations 1
A.2 A Summary of General Topology 3
A.3 Functional Analysis 4
B Probability Theory 5
B.1 Measure Spaces 5
B.2 Schwartz Distributions 10
B.3 Random Variables 11
C Random Processes 14
C.1 General Random Processes 14
C.2 Markov Chains 16
C.3 Ergodic Theory 18
C.4 Discrete Parameter Martingales 18
C.5 Continuous Martingales 20
C.6 Stochastic Integral 24
D Wiener–Lévy Calculus 26
D.1 Brownian Motion 26
D.2 Wiener Space 32

2. Computation of Expectations in Finite Dimension 34

A Mathematical Framework of Simulation 34
A.1 Narrow Convergence 34
A.2 Narrow Convergence and μ-Riemann Integrable Functions 35
A.3 Application to Simulation 38
A.4 Distribution Functions 40

A.5 Uniformly Distributed Sequences and Discrepancy 44
B The Monte Carlo Method 46
B.1 Algorithms for Simulation 46
B.2 Glivenko–Cantelli Theorem 48
B.3 Central Limit Theorem for Empirical Distributions 50
B.4 Stopping Criteria and Asymptotic Rate 54
B.5 Rate of Convergence of Stochastic Approximation Algorithms 58
C Low-Discrepancy Sequences 67
C.1 Koksma–Hlawka Inequality 67
C.2 Lower Bounds of Discrepancy 69
C.3 Irrational Translations on the Torus 71
C.4 Other Usual Sequences with Low Discrepancy 75
C.5 Low-Discrepancy Sequences and Simulation 82
C.6 Complement: Boreloid Sequences 86
D Numerical Computation of Conditional Expectation 95
D.1 Complements on Conditional Expectation 95
D.2 Polynomial Approximation 98
D.3 Point by Point Calculation 106
D.4 Comparison of Both Methods on an Example 110

3. Simulation of Random Processes 113

A Integration in Large or Infinite Dimensions 113
A.1 Changing Dimension 113
A.2 Simulation of a Non-Riemann Integrable Random Variable 114
A.3 Effective Computation of the Expectation of Random Variables in L^1 116
A.4 Integration of Functionals of Stochastic Processes 118
A.5 The Shift Method 121
A.6 Rate of Convergence for the Shift Method 125
A.7 On the Shift Method in Finite Dimension 130
A.8 Complement: Quasi-every Sequence Is Shift Uniformly Distributed 133
B Representations of Stationary Fields 142
B.1 Wiener Integral 143
B.2 Spectral Representation of Stationary (Wide Sense) Fields 147
B.3 Linear Transformations of Stationary Fields 150
B.4 Second-Order Linear Differential Equations 153

B.5 Multidimensional Processes and Fields 156
B.6 Real-Valued Fields 158
B.7 Strictly Stationary Fields 160
B.8 Space and Time Discretizations 167
B.9 Spectrum Discretization 172
B.10 Simulating Stationary Processes with Stationary Germs 173
C Markov Processes 181
C.1 Jump Processes 182
C.2 Iterative or Recursive Simulation 183
C.3 Simulation of a Brownian Motion Until Hitting a Hyperplane 189
D Processes with Stationary Independent Increments 198
D.1 Definition and First Properties 198
D.2 Convolution Semi-groups on $\mathbb{R}^d$ 200
D.3 Convolution Semi-Groups on $\mathbb{R}_+$ 202
D.4 Examples 203
D.5 Subordinators 205
D.6 Stable PSIIs 209
D.7 Application of the Subordination in Bochner Sense to Simulation 213
D.8 The Bernstein Monoid 219
E Point Processes 221
E.1 Setting Out the Problem 221
E.2 Setting Up the Models 221
E.3 Examples of Probability Distributions 224
E.4 Computation of Physical Quantities 227

4. Deterministic Resolution of Some Markovian Problems 230

A Elements in Markovian Potential Theory 230
A.1 Notation and Setting: Feller Processes 230
A.2 Fundamental Representations 235
A.3 Feynman–Kac Formula 237
B Balayage Algorithms 241
B.1 Poincaré's Balayage 241
B.2 The Harmonic Kernel of a Markov Chain 244
B.3 Numerical Balayage 246
C Reduced Function Algorithm 250
C.1 Reduced Function Algorithms with Respect to a Sub-Markov Kernel 250
C.2 The Filling Scheme and the Réduite Property 252

C.3 Optimal Stopping 253
C.4 Computation of Reduced Functions with Respect to Semi-groups 255
D The Carré du Champ Operator 262
D.1 The Extended Generator and Its Domain 263
D.2 Application to Hedging Strategies in Incomplete Markovian Markets 265

5. Stochastic Differential Equations and Brownian Functionals 269
A Lipschitzian Stochastic Differential Equations: Ito's Theorem 269
A.1 Notation and Lemmas 269
A.2 Lipschitz hypotheses 270
A.3 Ito's Theorem 270
A.4 Two Properties of the Solution 273
B Discretization of SDEs 274
B.1 Euler Scheme: Convergence and Rate Estimates 274
B.2 Example: Wiener Iterated Integrals 278
B.3 Example of a Multistep Scheme 279
B.4 Do Not Use the Shift Method Along the Sample Paths 283
C Irregularity of Brownian Functionals 284
C.1 Lack of Continuity of Stochastic Integrals 284
C.2 Non-Riemann-Integrable Stochastic Integrals 287
C.3 Other Topologies on Wiener Space 293
C.4 Importance of L^p Estimates 293
C.5 Sequential Integration by Picking Smooth Functions in the Wiener Space 294
D Simulatable Functionals 296
D.1 Simulatable Functionals on $\{0,1\}^{\mathbb{N}}$ 296
D.2 Simulatable Functionals on $[0,1]^{\mathbb{N}}$ 297
D.3 Regularity Properties of Simulatable Functionals 299
D.4 Simulatable Functionals on a General Probability Space 300
D.5 The Case of Wiener Space 302
E Symbolic Expansions of Solutions to SDEs 306
E.1 Ito–Stratonovitch transformation 306
E.2 Hu–Meyer Formulas 306
E.3 Isobe–Sato Formula 310
E.4 Expansion of the Solution of an SDE as a Series of Iterated Stochastic Integrals 314

F Application of the Shift Method to Multiple Wiener Integrals and to Solutions of SDEs 319
F.1 First Factorization of the Wiener Space 320
F.2 Functions of Solutions of Lipschitz SDEs 321
F.3 Functions of Multiple Wiener Integrals 323
F.4 Other Factorizations of Wiener Space 326

Notes **330**

References **337**

Index **353**

Preface

A striking phenomenon in the recent evolution of the sciences is that there is scarcely a field of application in which the theory of probability does not play a role. In economics, finance, traffic and transport, signal and image processing, genetics, environmental sciences, as well as for modeling structures exposed to wind, waves, or earthquakes, and for the science of materials—the standard models make use of random variables and stochastic processes, and sometimes the finer points of stochastic analysis.

Curiously, under the pretext that there is randomness in the model, engineers sometimes allow crude treatments that they would not accept in the analysis of deterministic models, as if random phenomena need not be treated with rigor. For instance, calculations with second-order moments and gaussian hypotheses are confused, extreme-value laws are mistakenly applied to nonadditive quantities, although the asymptotic estimates are not compatible with nonlinear changes of scale, the method of statistical linearization is used without the necessary precautions, asymptotic laws are used instead of the actual laws without estimating the difference between them, and so on. These misuses are encouraged by the current fashion for anything heuristic (because it resembles informal scientific investigation). The serious fault in these practices is that the results they give can neither usefully support a decision nor contribute to a discussion of the comparative risks of different choices, because there is no information about their accuracy.

The main motivation for this book is to give greater rigor to numerical treatments of stochastic models. This approach turns out to be scientifically fertile, since it extends the domain of application of mathematics and allows proofs that are an important preliminary to implementation on computers.

The quantities that we calculate in stochastic models are mathematical expectations (since probabilities of events, variances, and so on are

also expectations). If we forbid the crude methods outlined above, what methods are valid? First of all, in many cases we can compute expectations by analytic methods. It follows from the mathematical properties of the model that the desired expectations must satisfy certain deterministic equations, which we can solve numerically. Chapter 4, on Markovian problems, illustrates this situation. In other cases, we can use simulation methods, in which we pick points from a space equipped with a probability measure and take the average value over these points of the random variable whose expectation we want to calculate. These methods can be applied to a great variety of situations, and for nearly all the mathematically well-defined random variables (however, see the examples of non-Riemann integrable variables in Chapter 3, Section A and Chapter 5, Section C).

Simulation methods split into two families: those that use an imitation of randomness, the Monte Carlo method and the shift method (one based on the law of large numbers, the other on the pointwise ergodic theorem), are only absolutely rigorous if we suppose that the imitation of randomness is perfectly realized, which allows us to establish results by probabilistic reasoning. The other family contains the so-called quasi-Monte Carlo methods, in which we pick sequences of points, not randomly, but so that they have an asymptotic property of distribution that allows us to use them to calculate an expectation by averaging (provided we go about it correctly). These methods are rigorous and give demonstrable results, but they are less widely applicable than the Monte Carlo and shift methods. In particular, we do not know how to apply them effectively in infinite dimensions.

This book is based on a postgraduate course given by the authors at Paris 6 University. Like any work of applicable mathematics, it lies between the extreme points formed by a theoretical academic treatise, an elementary pedagogical account, and a specialized technical manual for engineers. Taking into account that computer systems that use both numbers and symbols and implement simulation methods will soon appear, it is important to decide precisely what knowledge in this field should be kept on paper. This has led us to accent the sorts of problems raised by new methods and to point out openings for research. The book nevertheless remains introductory and presents the current state of the main methods and ideas, the cases for which they have been proved, and the boundary areas where they work less well and need improvement.

We begin with a preliminary chapter in which we recall the main ideas from analysis and probability theory that are required in the sequel. Most of the notation is defined here. In principal, this chapter should not be read exhaustively; the reader can refer to it periodically in the course of later chapters.

In Chapter 2, we explain the mathematical framework in which the ideas of simulation, both for Monte Carlo and quasi-Monte Carlo methods, can be correctly expressed. The fundamental notions here are those of narrow convergence and of a μ-Riemann integrable function. We also define the discrepancy of a sequence and give equivalent conditions for uniform distribution. These fundamental ideas are applied in Section B to the Monte Carlo method. We have stated the results that are simplest and most useful for applications, and points of entry into the vast literature of deep results are given in the notes. We emphasize the importance of stopping conditions and asymptotic speeds and devote a section to showing, by explicit examples, the crucial importance of stopping conditions in algorithms for stochastic approximation. Section C is devoted to quasi-Monte Carlo methods. We give the Koksma–Hlawka inequality in a simplified form and investigate the main families of low-discrepancy sequences. We also explain the use of low-discrepancy sequences in simulation (which, given care, is straightforward) and give a detailed discussion of the difference in nature from the Monte Carlo method. We continue by suggesting avenues for research involving mixed methods lying between the Monte Carlo method and quasi-Monte Carlo methods. The chapter ends with a section on calculating conditional expectations where Monte Carlo and quasi-Monte Carlo methods are used for polynomial approximation and discretization. It is very surprising that this has not previously been covered in the literature, other than in specialized works on numerical filtering.

Chapter 3 deals with the simulation of the main families of random processes and fields. Since the canonical probability space on which we define a random process is most often infinite-dimensional, we must adopt a more precise approach than in Chapter 2; this will also be useful in Chapter 5 for the study of Wiener space. Despite appearances, the infinite-dimensional case is very frequent (most often, this means only that the dimension is not bounded, as when calculating the expected stopping time of a Markov chain), it is the appropriate domain of application for the shift method. Although based on classical theory (Birkhoff's pointwise ergodic theorem of 1931), it seems that the considerable advantages of this method, as far as computers are concerned, have only been popularized since the 1988 course at Paris 6 University, which developed into this book. The method and the principles for its implementation are accompanied here by results on the speed of convergence. There are many open problems, such as that of finding quick ways to estimate speed of convergence, particularly when the method is used in finite dimensions, where it can also be very useful. The section ends with an account of research on generic points for the shift. The rest of Chapter 3 is devoted to stationary processes and fields.

Markov processes, and point processes. Stationary processes, although less objects of research than in the past, remain a very useful tool for engineers. We give an account of them that does not require previous knowledge and present the simpler ideas concerning their simulation. We return to Markov processes in the next chapter; here we use little more than their definition. Central for applications in Paul Lévy's idea of recursive simulation. In our account of the simulation of Brownian motion before it reaches a hyperplane, we give an introduction to current research on conditional diffusion. Processes with independent increments are valuable tools for modeling that generalize models using Brownian motion; we take the opportunity to consider at the same time subordination in the sense of Bochner, which is rich in applications for simulation, but little known to engineers. The chapter ends with an introduction to point processes. These are marvelous tools for constructing models in very diverse fields of application.

Generally, Markovian models lend themselves nicely to analytic calculations where simulation methods should be used only as a last resort. Since we have given references to the classical works on Markov processes, we decided not to derive the more classical formulas, but rather to lay out the more useful tools for calculation without too much explanation. The sweeping algorithm, which is of considerable theoretical interest, is also an effective numerical tool, as we illustrate with an example. Its principle could be extended beyond the linear framework. The reduction algorithm is connected to a number of problems of optimal control and has recently found new applications for calculations involving "American" options in the mathematics of finance. It seemed to us to be of interest to introduce the carré du champ operator, which has a major theoretical role and yields useful formulas in certain applications, notably in finance.

The fifth and final chapter approaches questions of a slightly higher mathematical level. It shows that, thanks mainly to stochastic differential equations, the objects of stochastic calculus (i.e., defined in terms of Ito's stochastic integral in relation to Brownian motion) are both at the root of many models of physicists and engineers, and at the same time particularly difficult to simulate, because they are profoundly irregular. The treatment of stochastic differential equations and their discretization are standard; many improvements in this area can be found in the recent literature. However, it seemed important to give a detailed explanation of Wiener functionals, a field in which active theoretical research meets problems of application. We also introduce a tentative axiomatization of the notion of a simulatable random variable and its main properties. Although symbolic computation at present deals only partially with stochastic calculus, it seemed important for the future to give certain of the main formal

representations attached to stochastic integrals and stochastic differential equations (the formulas of Hu–Meyer and Isobe–Sato). In the last section we show how the shift method can be applied to Wiener space, not along the paths (which does not work), but by using representations as a series of independent variables, that is, on product spaces.

We must emphasize that there are questions in probability theory, related to numerical methods, that we do not address, or do not cover fully, in this book. Notable among those covered further in the literature are:

- classical methods in simulation (algorithms, reduction of variance, etc.)
- convergence of empirical processes
- stochastic approximation and stochastic control
- computation of expectation with respect to an invariant distribution of a recurrent Markov chain
- probabilistic methods for solving equations in physics
- fine methods for discretizing stochastic differential equations
- numerical aspects for nonlinear filtering

But this book is introductory, and its main point is the numerical use of the shift method, which will without doubt be of prolonged use in the areas mentioned above. We thank the researchers of the Centre d'Enseignement et de Recherche en Mathématiques Appliquées de l'Ecole des Ponts et Chaussées at Paris who have practiced and perfected the methods given in the book, as well as the works on which we have relied. We would also like to thank Alexander Scott and Philipp Regalia for their help in preparing the English version.

Nicolas Bouleau
Dominique Lépingle

Paris, France

Numerical Methods for Stochastic Processes

CHAPTER 1

Preliminaries

The theory and applications of stochastic processes are constantly growing and it is becoming ever more difficult to write a book on this topic which is simultaneously advanced, self-contained, and of reasonable length. Since we are concerned only with the simulation of such processes, we have attempted to strike a balance between what deserves careful exposition and what we assume to be more or less familiar to the reader.

For those who are not acquainted with what we have left out, we have gathered in this chapter the definitions and results that will be used in the sequel. We do not include proofs and we refer for complete exposition to the excellent recent books by Ikeda and Watanabe [1981], Dacunha-Castelle and Duflo [1982, 1983], Durrett [1984], Rogers and Williams [1987], and Revuz and Yor [1991]. In other words this chapter is best used for reference: take a quick glance through it and carry on, returning whenever you meet some notation, definition, or theorem which is handled without particular explanation.

A SET THEORY AND GENERAL TOPOLOGY

A.1 Some Notations

A.1.1

The set of real numbers is denoted by $\mathbb{R}$, and we write $\mathbb{R}_+ = [0, \infty[$, $\mathbb{R}_+^* =]0, \infty[$. For x, y in $\mathbb{R}$, we set

$$x \wedge y = \min(x, y)$$
$$x \vee y = \max(x, y)$$
$$x^+ = x \vee 0$$
$$x^- = (-x) \vee 0$$

$|x| = x^+ + x^-$

$[x]$ = integral part of x (largest integer not exceeding x)

$\{x\}$ = fractional part of x $(x - [x])$.

Similarly, for $x = (x_1, \ldots, x_d)$ and $y = (y_1, \ldots, y_d)$ in $\mathbb{R}^d$, we set

$$[x] = ([x_1], \ldots, [x_d])$$
$$\{x\} = (\{x_1\}, \ldots, \{x_d\})$$
$$x \wedge y = (x_1 \wedge y_1, \ldots, x_d \wedge y_d)$$
$$\langle x, y \rangle = x_1 y_1 + \cdots + x_d y_d$$
$$|x| = \langle x, x \rangle^{1/2}.$$

If $x \wedge y = x$, we define

$$]]x, y]] =]x_1, y_1] \times \cdots \times]x_d, y_d]$$
$$[[x, y]] = [x_1, y_1] \times \cdots \times [x_d, y_d].$$

The notation $]]x, y]]$ will also be used when some of the x_i's are $-\infty$. We denote by x^t the transpose of the vector x and identify xy^t with $\langle x, y \rangle$.

A.1.2.
A real number r is called an *algebraic number* if it is a solution to some polynomial equation

$$a_0 + a_1 x + \cdots + a_m x^m = 0$$

with integral coefficients $(a_i; i = 0, \ldots, m)$.

A.1.3
A *Gaussian integer* is a complex number $a + ib$, where a and b belong to $\mathbf{Z}$. The ring of Gaussian integers is denoted by $\mathbf{Z}[i]$.

A.1.4
If B is a subset of A, the *difference* of A and B is denoted by $A \backslash B$, or more simply B^c if there is no doubt about A. If A and B are disjoint sets, we will sometimes write $A + B$ instead of $A \cup B$. For instance, the *symmetric difference* of the sets A and B is

$$A \triangle B = (A \cap B^c) + (B \cap A^c).$$

To every subset B of A we assign a function 1_B, to be called the *indicator* of B, defined by

$$1_B(x) = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{if } x \notin B. \end{cases}$$

A.2 A Summary of General Topology

A.2.1

Let $(E, \mathscr{T})$ be a topological space. If A is a subset of E, we write $\mathring{A}$ for the *interior* of A, $\overline{A}$ for the *closure* of A, and ∂A for the *boundary* of A.

A.2.2

Given the topological space $(E, \mathscr{T})$, it is natural to ask whether or not there exists a metric d on E that induces the topology $\mathscr{T}$. The topological space $(E, \mathscr{T})$ is said to be *metrizable* if such a metric exists. A metric space (E, d) is *complete* if every Cauchy sequence in E converges to a point in E. A topological space $(E, \mathscr{T})$ is a *Polish space* if it is separable and if there exists a metric d that induces the topology and makes E a complete metric space.

A.2.3

A normed vector space is a *Banach space* if the induced metric makes it a complete metric space. A useful theorem, for both finite- and infinite-dimensional real normed vector spaces, is the *Hahn–Banach theorem*: Let E be a vector space and F be a subspace of E. Suppose $\lambda: F \to \mathbb{R}$ is linear and there exists $c > 0$ such that $|\lambda(x)| \leq c|x|$ for all x in F. Then there exists an extension of λ to a linear mapping from E to $\mathbb{R}$ with the same bound c.

A.2.4

Let E be a real or complex vector space and let $(x, y) \rightsquigarrow \langle x, y \rangle$ be a symmetric or Hermitian form on E. This form is *positive definite* if $\langle x, x \rangle > 0$ for all $x \neq 0$. A real or complex *Hilbert space* is a vector space with a positive definite symmetric or Hermitian form, which is complete under the corresponding norm

$$|x| = \langle x, x \rangle^{1/2}.$$

Let $(v_i, i \in I)$ be a family of elements of a Hilbert space H such that $v_i \neq 0$ for all i. We can take the subspace of all linear combinations

$$c_1 v_{i_1} + \cdots + c_n v_{i_n}$$

for finite subfamilies $(v_{i_j}; j = 1, \ldots, n)$ of (v_i) and real or complex coefficients (c_j). We say that the family (v_i) is *total* in H if the closure of this subspace is equal to H.

A.2.5
Let C be a nonempty closed convex set in a Hilbert space H. For every $x \in H$, there exists a unique $p(x) \in C$ such that

$$|x - p(x)| = \inf\{|x - y|; y \in C\}.$$

We call $p(x)$ the *orthogonal projection* of x onto C. In particular, every non-empty closed convex set in a Hilbert space contains a unique element of smallest norm.

A.2.6
A topological space $(E, \mathscr{T})$ is a *Hausdorff space* if for every pair of distinct points a, b in E there are disjoint open sets U and V such that a is in U and b is in V. We say that E is *locally compact* if every point in E has a compact neighborhood. A subset of E is said to be *relatively compact* if its closure is compact.

A.2.7

For $s \in \mathbf{N}^*$ the Euclidean space $\mathbb{R}^s$ is a real Hilbert space with the usual symmetric form $\langle x, y \rangle$. The induced topology makes it a locally compact Hausdorff space. The *quotient topology* on the *torus* $\mathbf{T}_s = \mathbb{R}^s/\mathbf{Z}^s$ is the largest topology such that the canonical projection $\pi_s \colon \mathbb{R}^s \to \mathbb{R}^s/\mathbf{Z}^s$ is continuous. This topology makes $\mathbf{T}_s$ a metrizable compact Hausdorff space.

A.3 Functional Analysis

A.3.1
A real or complex function f on a locally compact Hausdorff space E is said to *vanish at infinity* if, for every $\epsilon > 0$ there exists a compact set $K \subset E$ such that $|f(x)| < \epsilon$ for every x not in K. We denote by $\mathscr{C}_0(E)$ the class of all continuous functions on E that vanish at infinity.

A.3.2
Let $(E, \mathscr{T})$ be a topological space. We denote by $\mathscr{C}(E)$ (resp., $\mathscr{C}(E; \mathbb{R}^d)$, $\mathscr{C}_b(E)$, $\mathscr{C}_b(E; \mathbb{R}^d)$) the set of real-valued continuous functions on E (resp.,

$\mathbb{R}^d$-valued continuous functions, bounded continuous real-valued functions, bounded continuous $\mathbb{R}^d$-valued functions).

A.3.3
An *algebra* A of real-valued functions on E is a ring of functions such that if $f, g \in A$ then $f + g \in A, fg \in A$, and $cf \in A$ for every $c \in \mathbb{R}$. If E is compact, then every continuous function on E is bounded and we have the following *Stone-Weierstrass theorem*: If A is a subalgebra of $\mathscr{C}(E)$ which separates points and contains the constant functions, then the uniform closure of A is equal to $\mathscr{C}(E)$. There is also a complex version of the Stone-Weierstrass theorem for self-conjugate subalgebras (i.e. if $f \in A$, then $\bar{f} \in A$, where $\bar{f}$ is defined by $\bar{f}(x) = \overline{f(x)}$). If E is a compact interval $[a, b]$, we conclude that every continuous real function on $[a, b]$ can be uniformly approximated by polynomials on $[a, b]$: this is the so-called *Weierstrass theorem.*

A.3.4
It is sometimes useful to have a criterion for when a subset of a function space is relatively compact. This is the purpose of *Ascoli's theorem*: Let (E, d) be a compact metric space, let F be a Banach space, and let Φ be a subset of the space $\mathscr{C}(E; F)$ of continuous maps from E into F. Suppose that for every $\epsilon > 0$ there exists $\delta > 0$ such that whenever $x, y \in E$ and $d(x, y) < \delta$, then $|f(x) - f(y)| < \epsilon$ for all $f \in \Phi$. Suppose also that for every $x \in E$, the set

$$\phi(x) = \{f(x); f \in \Phi\}$$

is relatively compact. Then Φ is relatively compact in $\mathscr{C}(E; F)$ endowed with the sup norm.

B PROBABILITY THEORY

B.1 Measure Spaces

B.1.1
If $\mathscr{A}$ is a σ-field (σ-algebra) in a set Ω, then $(\Omega, \mathscr{A})$ is called a *measurable* space and the members of $\mathscr{A}$ are called the *measurable sets.* We shall most often deal with topological spaces $(E, \mathscr{T})$ and there is a smallest σ-field $\mathscr{B}$ in E such that every open set in E belongs to $\mathscr{B}$. The members of $\mathscr{B}$ are called the *Borel sets* of E and $\mathscr{B}$ is called the *Borel* σ-field in E. We often denote this σ-field by $\mathscr{B}(E)$ or $\mathscr{E}$. A mapping $f : E \to \mathbb{R}$ or $\mathbf{C}$ such that $f^{-1}(V)$ is a Borel set for every open set V in $\mathbb{R}$ or $\mathbf{C}$ will be

called a *Borel function*. The set of bounded Borel functions on E will be denoted by $b\mathscr{E}$.

B.1.2
Let $(\Omega, \mathscr{A})$ and $(\Omega', \mathscr{A}')$ be two measurable spaces. A function $f : \Omega \to \Omega'$ is called *measurable* if $f^{-1}(\mathscr{A}') \subset \mathscr{A}$. If $\mathscr{F} = (f_i, i \in I)$ is a set of functions from Ω into measurable spaces $((\Omega_i, \mathscr{A}_i), i \in I)$, then $\sigma(\mathscr{F})$ is the smallest σ-field in Ω which makes every f_i a measurable function. In the same way, if $\mathscr{C}$ is a collection of subsets of Ω, then $\sigma(\mathscr{C})$ is the σ-field *generated* by $\mathscr{C}$, that is, the smallest σ-field containing $\mathscr{C}$.

B.1.3
A collection $\mathscr{C}$ of subsets of Ω is an *algebra* if it contains Ω and is closed under the set operations $(\cap, ^c)$. A collection $\mathscr{U}$ of subsets of Ω is a *monotone class* if

$$\{A_n\} \subset \mathscr{U} \text{ and } A_1 \subset A_2 \subset \ldots \text{ imply } \cup A_n \in \mathscr{U}$$

and

$$\{A_n\} \subset \mathscr{U} \text{ and } A_1 \supset A_2 \supset \ldots \text{ imply } \cap A_n \in \mathscr{U}.$$

If $\mathscr{C}$ is an algebra and $\mathscr{U}$ is a monotone class with $\mathscr{C} \subset \mathscr{U}$, then $\sigma(\mathscr{C}) \subset \mathscr{U}$. This is the *monotone class theorem* for sets. There is also a monotone class theorem for functions: Let $\mathscr{H}$ be a vector space of functions from Ω to $\mathbb{R}$ such that

(i) $1 \in \mathscr{H}$,

(ii) if $(h_n, n \in \mathbf{N})$ is a non-decreasing sequence of non negative elements of $\mathscr{H}$ with $\lim_n h_n = h$, where h is finite (resp., bounded), then $h \in \mathscr{H}$.

Suppose that $\mathscr{H}$ contains the indicator function 1_A of every set A in some collection $\mathscr{C}$ of subsets of Ω, which is closed under finite intersections. Then $\mathscr{H}$ contains every real-valued (resp., bounded real-valued) function on Ω which is $\sigma(\mathscr{C})$-measurable.

B.1.4
A (positive) *measure* is a function μ from a σ-field $\mathscr{A}$ to $[0, \infty]$, which is countably additive and has $\mu(\varnothing) = 0$. The space $(\Omega, \mathscr{A}, \mu)$ is then called a *measure space*. A measure μ on $(\Omega, \mathscr{A})$ is said to be *finite* if

$\|\mu\| \equiv \mu(\Omega) < \infty$ and σ-*finite* if Ω is a countable union of measurable A_i with $\mu(A_i) < \infty$.

B.1.5
An important role is played by the *Caratheodory extension theorem*: Let $\mathscr{C}$ be an algebra in Ω and let $\mu : \mathscr{C} \to [0, \infty]$ be countably additive and σ-finite with $\mu(\varnothing) = 0$. Then μ can be uniquely extended to a measure on $\sigma(\mathscr{C})$.

B.1.6
Given a measure space $(\Omega, \mathscr{A}, \mu)$, we may associate with any A in $\mathscr{A}$ another measure space $(A, \mathscr{A} \cap A, \mu|_A)$, where

$$\mu|_A(B \cap A) = \mu(B \cap A), \qquad B \in \mathscr{A};$$

$\mu|_A$ is called the *restriction* of μ to A.

B.1.7
Let $f : (\Omega, \mathscr{A}) \to (\Omega', \mathscr{A}')$ be a measurable function. If μ is a measure on $(\Omega, \mathscr{A})$, then $\nu = \mu \circ f^{-1}$ is a measure on $(\Omega', \mathscr{A}')$, which we denote by $f_*\mu$.

B.1.8
Suppose that, for every i in some (possibly uncountable) index set I, $(\Omega^i, \mathscr{A}^i)$ is a measurable space. Let $\Omega^{(I)}$ denote the Cartesian product

$$\Omega^{(I)} = \prod_{i \in I} \Omega^i.$$

The *product* σ-*field* $\mathscr{A}^{(I)}$ is defined to be the σ-field $\sigma(\pi^i, i \in I)$, where $\pi^i : \Omega^{(I)} \to \Omega^i$ are the canonical projections. If $I = \{1, 2\}$, the product σ-field is denoted by $\mathscr{A}^1 \otimes \mathscr{A}^2$.

If μ_1 and μ_2 are σ-finite measures on $(\Omega^1, \mathscr{A}^1)$ and $(\Omega^2, \mathscr{A}^2)$, respectively, then there exists a unique σ-finite measure on $(\Omega^1 \times \Omega^2, \mathscr{A}^1 \otimes \mathscr{A}^2)$ determined by the relation

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2), \qquad A_i \in \mathscr{A}^i.$$

This is called the *product measure* $\mu_1 \otimes \mu_2$.

B.1.9
The Lebesgue measure in $[0, 1]$, $[0, 1[$, $\mathbb{R}_+$ or $\mathbb{R}$ is denoted by dx or λ_1. In the product spaces $[0, 1]^s$, $\mathbb{R}^s$, or $\mathbf{T}_s = \mathbb{R}^s/\mathbb{Z}^s$, for $s \in \mathbf{N}^*$, we shall write λ_s or $(\lambda_1)^s$ or dx or $(dx)^s$. In $[0, 1]^{\mathbf{N}} = [0, 1]^\infty$, the infinite-dimensional Lebesgue measure will be denoted by λ_∞.

B.1.10
Let $x \in \Omega$ and suppose that the subset $\{x\}$ belongs to a σ-field $\mathscr{A}$ on Ω. The *Dirac measure* δ_x on $(\Omega, \mathscr{A})$ is defined by

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

B.1.11
Let $(\Omega, \mathscr{A}, \mu)$ be a measure space. Two measurable real or complex functions f and g on $(\Omega, \mathscr{A})$ are said to be *equivalent* if $f(x) = g(x)$ for μ-almost every x. If $1 \leq p < \infty$, $L^p(\Omega, \mathscr{A}, \mu)$ is the space of equivalence classes of functions which satisfy

$$\|f\|_p = \left(\int_\Omega |f(x)|^p \, d\mu(x) \right)^{1/p} < \infty.$$

If $p = \infty$, we set

$$\|f\|_\infty = \inf \{K : \mu(\{x : |f(x)| > K\}) = 0\}.$$

The L^p-spaces ($1 \leq p < \infty$) are Banach spaces and the L^2-spaces are Hilbert spaces with the Hermitian form

$$\langle f, g \rangle = \int f(x) \overline{g(x)} \, d\mu(x).$$

If Ω is $\mathbb{R}$ or any subinterval of $\mathbb{R}$ and μ is Lebesgue measure, then the (equivalent classes of) continuous functions on Ω with compact support are dense in every L^p, $1 \leq p < \infty$. This is also true if $\Omega = \mathbb{R}^s$ and $\mu = \lambda_s$, for $s \in \mathbf{N}^*$.

B.1.12
If $(E, \mathcal{T})$ is a topological space with Borel σ-field $\mathcal{E}$, then a measure μ on $(E, \mathcal{E})$ is called a *Borel measure*. Let $(E, \mathcal{E})$ be a locally compact Hausdorff space and denote by $\mathcal{C}_c(E)$ the vector space of real-valued continuous functions on E with compact support (i.e. vanishing outside a compact set). Let $\Lambda : \mathcal{C}_c(E) \to \mathbb{R}$ be a positive linear map ($\Lambda f \geq 0$ whenever $f \geq 0$). Then (*Riesz representation theorem*) there exists a unique Borel measure on $(E, \mathcal{E})$ satisfying conditions (i) and (ii) below:

(i) If V is open,

$$\mu(V) = \sup\{\Lambda g; g \in \mathcal{C}(E), 0 \leq g \leq 1, \operatorname{supp}(g) \subset V\};$$

(ii) If A is a Borel set,

$$\mu(A) = \inf\{\mu(V); V \text{ open}, A \subset V\}.$$

Such a Borel measure is called a Radon measure and, for every f in $\mathcal{C}_c(E)$, we have

$$\Lambda f = \int f \, d\mu.$$

Now let $\Lambda : \mathcal{C}_c(E) \to \mathbb{R}$ be a linear map which is also continuous for the sup norm. It can be written as the difference of two continuous positive linear maps, so there exist two finite Radon measures μ_1, μ_2 such that

$$\Lambda f = \int f \, d\mu_1 - \int f \, d\mu_2.$$

The countably additive mapping $\mu : \mathcal{E} \to \mathbb{R}$ defined by $\mu = \mu_1 - \mu_2$ depends only on Λ and is called the *bounded measure* associated with Λ.

B.1.13
A function f on $\mathbb{R}^d$ is *nonnegative definite* if for any finite sets $\{x_i, i \in I\}$ in $\mathbb{R}^d$ and $\{\alpha_i, i \in I\}$ in $\mathbf{C}$,

$$\sum_{i,j \in I} f(x_i - x_j)\alpha_i \overline{\alpha_j} \geq 0.$$

If, moreover, f is a continuous function, then (*Bochner's theorem*) there exists a finite measure μ on $\mathbb{R}^d$ such that

$$f(x) = \int e^{i\langle \lambda, x\rangle} \mu(d\lambda), \qquad x \in \mathbb{R}^d.$$

In the discrete case, we have the *Herglotz lemma*: Suppose $f : \mathbf{Z} \to \mathbf{C}$ satisfies

$$\sum_{n,m \in I} f(n-m)\alpha_n \overline{\alpha_m} \geq 0$$

for any finite subset I of $\mathbf{Z}$ and any set $\{\alpha_n, n \in I\}$ in $\mathbf{C}$; then there exists a finite measure μ on $]-\pi, +\pi]$ such that

$$f(n) = \int e^{i\lambda n} \mu(d\lambda), \qquad n \in \mathbf{Z}.$$

B.2 Schwartz Distributions

B.2.1

Let $D_i = \partial/\partial x_i$ be the ith partial derivative applied to functions on $\mathbb{R}^s$. For a p-triple $(p_1, \ldots, p_s) = p$ of integers ≥ 0, we let

$$D^p = D_1^{p_1} \cdots D_s^{p_s} \qquad \text{and} \qquad |p| = p_1 + \cdots + p_s.$$

The space $\mathscr{D}$ (or $\mathscr{C}_c^\infty(\mathbb{R}^s)$) of all infinitely differentiable real-valued functions on $\mathbb{R}^s$ with compact support is called the space of test functions. A linear map $T : \mathscr{D} \to \mathbb{R}$ is a *distribution* if for every compact set K in $\mathbb{R}^s$, there exists a constant A_K and an integer m such that

$$|T\varphi| \leq A_K \sup_{|p| \leq m} \sup_{x \in K} |D^p \varphi(x)|$$

for every φ in $\mathscr{D}$ that vanishes outside K.

B.2.2

If f is a Borel function on $\mathbb{R}^s$, which is λ_s-integrable on every compact subset of $\mathbb{R}^s (f \in \mathscr{L}^1_{loc}(\mathbb{R}^s))$, then the map

$$T_f(\varphi) = \int f\varphi d\lambda_s, \qquad \varphi \in \mathscr{D}$$

is a distribution. If μ is a Radon measure or a bounded measure on $\mathbb{R}^s$ then the map

$$T_\mu(\varphi) = \int \varphi d\mu, \qquad \varphi \in \mathscr{D}$$

is a distribution. If T is a distribution, we define a new distribution D_iT, for $i = 1, \ldots, s$, by putting

$$(D_iT)(\varphi) = -T(D_i\varphi), \qquad \varphi \in \mathscr{D}.$$

This notation is consistent with the ordinary derivation of functions since, if f is a $\mathscr{C}^1$-function, we have

$$D_iT_f = T_{D_if}.$$

We may also define D^pT for any differential operator D^p in a natural way.

B.3 Random Variables

B.3.1

We shall mainly be interested in finite measures with total mass equal to 1. Such a measure is called a *probability measure* and the set of all probability measures on $(\Omega, \mathscr{A})$ is denoted by $\mathscr{M}_+^1(\Omega, \mathscr{A})$. If $\mathbb{P} \in \mathscr{M}_+^1(\Omega, \mathscr{A})$, the triple $(\Omega, \mathscr{A}, \mathbb{P})$ is called a *probability space*.

B.3.2

A *random element* is a measurable function X from a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ into a measurable space $(\Omega', \mathscr{A}')$. If $(\Omega', \mathscr{A}') = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ (resp., $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$), it is called a *random variable* (resp., a d-dimensional random variable). The *distribution*, or *law*, of the random element X is the probability measure $\mathbb{P} \circ X^{-1} = X_*\mathbb{P}$. The distribution function F_X of a random variable X is defined by

$$F_X(x) = \mathbb{P}(X \in]-\infty, x]), \qquad x \in \mathbb{R}.$$

We shall also use the *distribution function* of an $\mathbb{R}^d$-valued random variable:

$$F_X(x) = \mathbb{P}(X \in]]-\infty, x]]), \qquad x \in \mathbb{R}^d.$$

A *density function* of a d-dimensional random variable (or of its distribution) is any function f such that, for every Borel set B in $\mathbb{R}^d$,

$$\mathbb{P}(X \in B) = \int_B f(x)\,dx.$$

B.3.3
Let X be a $\mathbb{P}$-integrable d-dimensional random variable. The *expectation* or *mean value* of X is defined by

$$\mathbb{E}X = \int X\,dP.$$

This depends only on the equivalence class of X in $L^1(\Omega, \mathscr{A}, \mathbb{P}; \mathbb{R}^d)$. If X and Y are two (equivalence classes of) random variables in $L^2(\Omega, \mathscr{A}, \mathbb{P})$, their *covariance* is

$$\mathrm{Cov}\,(X, Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y.$$

A d-dimensional random variable X is *centered* (has zero mean) if its components are $\mathbb{P}$-integrable and $\mathbb{E}X = 0$.

B.3.4
If X is a d-dimensional random variable with distribution μ_X, the *characteristic function* of X (or of μ_X) is the function $\varphi_X : \mathbb{R}^d \to \mathbf{C}$ defined by

$$\varphi_X(u) = \mathbb{E}(e^{i\langle u, X\rangle}) = \int \exp\{i\langle u, x\rangle\}\mu_X(dx).$$

The function φ_X uniquely determines μ_X.

B.3.5
A *real Gaussian random variable* X is a random variable with distribution $N(m, v)$ determined by the characteristic function

$$\varphi_X(u) = \exp\{imu - \tfrac{1}{2}vu^2\}, \qquad u \in \mathbb{R},$$

where $m \in \mathbb{R}$, $v \in \mathbb{R}_+$. A d-dimensional Gaussian random variable is a d-dimensional random variable such that any linear combination of its components is a real Gaussian random variable. Its distribution is denoted by $N_d(m, \Sigma)$, where m is the mean vector and Σ the covariance matrix of the components.

B.3.6

A collection of σ-fields $(\mathcal{F}_i, i \in I)$ contained in $\mathcal{A}$ is said to be independent if, for any finite subset J of I and any $(A_j \in \mathcal{F}_j, j \in J)$,

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j).$$

Random elements $(X_i, i \in I)$ on $(\Omega, \mathcal{A}, \mathbb{P})$ are said to be independent if the σ-fields $(\sigma(X_i), i \in I)$ are independent.

B.3.7

A sequence of d-dimensional random variables $(X_i, i \in I \subset \mathbf{N})$ is said to be i.i.d. (independent, identically distributed) with common law μ if the X_i are independent and have the same distribution μ. Such sequences satisfy the *strong law of large numbers*: if $\int |x| \mu(dx) < \infty$, then

$$n^{-1} \sum_{i=1}^{n} X_i \to \int x \mu(dx) \qquad \mathbb{P} - \text{a.s.}$$

B.3.8

A sequence of d-dimensional random variables $(Y_n, n \geq 1)$ *converges in distribution* or *in law*, to the $\mathbb{R}^d$-valued random variable Y (resp., to the probability measure μ on $\mathbb{R}^d$) if for every f in $\mathcal{C}_b(\mathbb{R}^d)$,

$$\mathbb{E}[f(Y_n)] \to \mathbb{E}[f(Y)]$$

(resp., $\mathbb{E}[f(Y_n)] \to \int f(x) \mu(dx)$).

We shall write $Y_n \xrightarrow{\mathcal{L}} Y$ (resp., $Y_n \xrightarrow{\mathcal{L}} \mu$). Notice that each Y_n may be carried by its own probability space. A basic fact in probability theory is the *central limit theorem*: Let $(X_n, n \geq 1)$ be an i.i.d. sequence of d-dimensional random variables with common law μ such that

$$\int |x|^2 \mu(dx) < \infty.$$

Set

$$m = \int x\mu(dx)$$

$$D = \int (x - m)^t (x - m)\mu(dx)$$

$$S_n = \sum_{i=1}^{n} X_i \qquad n \geq 1.$$

Then $n^{-1/2}(S_n - nm)$ converges in distribution to $N_d(0, D)$.

B.3.9
Let $\mathscr{B}$ be a sub-σ-field of $\mathscr{A}$. The *conditional expectation* given $\mathscr{B}$ of a random variable X in $L^1(\Omega, \mathscr{A}, \mathbb{P})$ is the random variable (up to a $\mathbb{P}$-equivalence class) $\mathbb{E}(X|\mathscr{B})$ defined by the two following properties:

(i) $\mathbb{E}(X|\mathscr{B})$ is $\mathscr{B}$-measurable,
(ii) for every $B \in \mathscr{B}$,

$$\int_B X d\mathbb{P} = \int_B \mathbb{E}(X|\mathscr{B}) d\mathbb{P}.$$

For $A \in \mathscr{A}$, we define the *conditional probability of A given* $\mathscr{B}$ by the equation

$$\mathbb{P}(A|\mathscr{B}) = \mathbb{E}(1_A|\mathscr{B}) \qquad \mathbb{P} - \text{a.s.}$$

C RANDOM PROCESSES

C.1 General Random Processes

C.1.1
We shall most often be interested in infinite-dimensional situations, where there is an infinite collection of random variables defined on the same probability space. This collection will be called a *random field* or *random*

process. The index set of these collections will be either $\mathbb{R}^d$ or $\mathbb{R}$ or $\mathbf{Z}$, or most often $\mathbb{R}+$ or $\mathbf{N}$.

C.1.2

Let us begin by stating *Kolmogorov's extension theorem* for product spaces, which is the main tool for constructing random fields with given finite-dimensional distributions. Let I be an infinite index set and for each $i \in I$ let E^i be a Polish space with Borel σ-field $\mathscr{E}^i$. For sets $F \subset I$ we denote by $E^{(F)}$ the product space $\prod_{i\in F} E^i$ and by $\mathscr{E}^{(F)}$ the product σ-field $\prod_{i\in F} \mathscr{E}^i$. For sets $I \supset G \supset F \neq \varnothing$, let π_F^G denote the canonical projection from $E^{(G)}$ onto $E^{(F)}$. Suppose that for every finite set F we are given a probability measure $\mathbb{P}_F$ on $(E^{(F)}, \mathscr{E}^{(F)})$ such that for any two finite sets $G \supset F \neq \varnothing$, $\mathbb{P}_F = \pi_{F*}^G \mathbb{P}_G$. Then there exists a unique probability measure $\mathbb{P}$ on $(E^{(I)}, \mathscr{E}^{(I)})$ such that $\mathbb{P}_F = \pi_{F*}^I \mathbb{P}$ for all finite $F \neq \varnothing$.

C.1.3

Let $(E, \mathscr{E})$ be a Polish space with its Borel σ-field and let I be a set. We write E^I for the product space $\prod_{i\in I} E^i$, where $E^i = E$ for all $i \in I$, and $\mathscr{E}^{\otimes I}$ for the related product σ-field. A random process X with state space $(E, \mathscr{E})$ and parameter set I is a collection

$$X = (X_i; i \in I)$$

of measurable functions from a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ to $(E, \mathscr{E})$. We can view X globally as a measurable map

$$X : (\Omega, \mathscr{A}, \mathbb{P}) \to (E^I, \mathscr{E}^{\otimes I}).$$

Let $\mathscr{L}_X = X_* \mathbb{P}$ be the law of X. In this setup the canonical process associated with X is the process of canonical projections $(Y_i; i \in I)$ defined on $(E^I, \mathscr{E}^{\otimes I}, \mathscr{L}_X)$ by

$$Y_i(w) = w(i).$$

C.1.4

If $X = (X_i; i \in I)$ and $Z = (Z_i; i \in I)$ are processes with the same parameter set I, the same state space $(E, \mathscr{E})$ and same probability space $(\Omega, \mathscr{A}, \mathbb{P})$, we say that X and Z are *modifications* of each other if

$$\forall i \in I, \qquad \mathbb{P}(X_i = Z_i) = 1.$$

However, the right concept of two random processes being the same is the following one: X and Z are *indistinguishable* if

$$\mathbb{P}(\forall i \in I, X_i = Z_i) = 1.$$

A subset A of $I \times \Omega$ is said to be *evanescent* if 1_A is indistinguishable from the zero process.

C.2 Markov Chains

C.2.1

Let $(\Omega, \mathscr{F})$ and $(E, \mathscr{E})$ be measurable spaces. A *Markov kernel* from $(\Omega, \mathscr{F})$ into $(E, \mathscr{E})$ is a mapping $N : \Omega \times \mathscr{E} \to [0, 1]$ that satisfies the following properties:

(i) for fixed $\omega \in \Omega$, $C \rightsquigarrow N(\omega, C)$ is a probability measure on $(E, \mathscr{E})$;
(ii) for fixed $C \in \mathscr{E}$, $\omega \rightsquigarrow N(\omega, C)$ is $\mathscr{F}$-measurable.

To define a *sub-Markov kernel* we replace (i) by (i′):

(i′) for fixed $\omega \in \Omega$, $C \rightsquigarrow N(\omega, C)$ is a measure on $(E, \mathscr{E})$.

C.2.2

If $(\Omega, \mathscr{F}) = (E, \mathscr{E})$, a Markov (resp. sub-Markov) kernel from $(E, \mathscr{E})$ into itself is called a Markov (resp. sub-Markov) kernel on $(E, \mathscr{E})$. When dealing with a sub-Markov kernel P, it is convenient to adjoin an extra point δ to E, to produce an extended space $E' = E \cup \{\delta\}$ with σ-field $\mathscr{E}' = \mathscr{E} \cup \{B \cup \{\delta\} : B \in \mathscr{E}\}$. We set

$$P(x, \{\delta\}) = 1 - P(x, E) \qquad x \in E$$
$$P(\delta, A) = 1_A(\delta) \qquad A \in \mathscr{E}'$$

and the sub-Markov kernel P extends to a Markov kernel, which we also denote by P. The extra point δ is called a *coffin state* for P.

C.2.3

Let P be a sub-Markov kernel on $(E, \mathscr{E})$ and let π be a probability measure on $(E', \mathscr{E}')$. A sequence of random variables $X = (X_n, n \geq 0)$ taking values in E' is called a *Markov chain* with *transition function* P and *initial measure* π if

(i) the distribution of X_0 is π;
(ii) for $n \geq 1$ and $B \in \mathscr{E}'$,

$$\mathbb{P}(X_{n+1} \in B | \sigma(X_0, \ldots, X_n)) = \mathbb{P}(X_{n+1} \in B | \sigma(X_n)) = P(X_n, B) \qquad \mathbb{P} - \text{a.s.}$$

Therefore all finite-dimensional distributions of X are uniquely determined.

Now suppose that $(E, \mathscr{E})$ is a Polish space with its Borel σ-field. We may use Kolmogorov's extension theorem to get, for any sub-Markov kernel P and any probability measure π on $(E', \mathscr{E}')$, a probability measure $\mathbb{P}$ on $((E')^{\mathbf{N}}, (\mathscr{E}')^{\otimes \mathbf{N}})$ such that the coordinate mappings $(X_n, n \geq 0)$ are a Markov chain with transition function P and initial measure π. If P is fixed, this probability measure is denoted by $\mathbb{P}_\pi$; we will write $\mathbb{P}_x$ when π is the Dirac measure δ_x. The random time $\zeta = \inf\{n \geq 0 : X_n = \delta\}$ is called the *lifetime* of X, and

$$\mathbb{P}(\omega : X_n(\omega) = \delta, \forall n \geq \zeta(\omega)) = 1.$$

C.2.4

In the setup defined above, we define the *shift operator* θ on $\Omega = (E')^{\mathbf{N}}$ by the equation

$$\theta(x_0, x_1, \ldots) = (x_1, x_2, \ldots);$$

for every random variable η, we define the random variables $\theta^n \eta, n \geq 0$ by putting

$$(\theta^n \eta)(\omega) = \eta(\theta^n(\omega)).$$

The *Markov property* of the chain X can be given the following form: for every random variable η, every $n \geq 0$, and every $B \in \mathscr{E}'$,

$$\mathbb{P}_\pi(\theta^n \eta \in B | \sigma(X_0, \ldots, X_n)) = \mathbb{P}_{X_n}(\eta \in B) \qquad \mathbb{P}_\pi - \text{a.s.}$$

where the right-hand side is the function $\mathbb{P}_x(\eta \in B)$ evaluated at $x = X_n$.

C.3 Ergodic Theory

C.3.1
Let $(\Omega, \mathscr{E}, \mathbb{P})$ be a probability space. A measurable function $\theta : (\Omega, \mathscr{E}) \to (\Omega, \mathscr{E})$ is called a *measure-preserving transformation* if $\mathbb{P}(\theta^{-1}A) = \mathbb{P}(A)$ for every A in $\mathscr{E}$. A set $A \in \mathscr{E}$ is *almost invariant* for θ if $\mathbb{P}(A \triangle \theta^{-1}A) = 0$. A measure-preserving transformation is called *ergodic* if for every almost invariant set A we have $\mathbb{P}(A) = 0$ or 1. It is called *mixing* if for every A and $B \in \mathscr{E}$,

$$\lim_{n\to\infty} \mathbb{P}(A \cap \theta^{-n}B) = \mathbb{P}(A)\mathbb{P}(B).$$

Every mixing transformation is ergodic.

C.3.2
Let θ be an ergodic transformation and let ξ be a random variable in $L^1(\Omega, \mathscr{E}, \mathbb{P})$. Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \omega) = \mathbb{E}\xi \qquad \mathbb{P}-\text{a.s.}$$

This is *Birkhoff's pointwise ergodic theorem.*

C.3.3
Let $(\Omega, \mathscr{E}, \mathbb{P}, \theta)$ be a Polish space with its Borel σ-field, a probability measure $\mathbb{P}$ and a one-to-one measure-preserving transformation θ. Suppose also that the set of points ω in Ω such that there exists an integer $n \geq 1$ with $\theta^n\omega = \omega$ has measure zero. Then (*Rohlin–Halmos lemma*), for every $\epsilon > 0$ and every positive integer n there exists a set $A \in \mathscr{E}$ such that

(i) the sets $A, \theta A, \ldots, \theta^{n-1}A$ are disjoint;
(ii) $\mathbb{P}(\cup_{i=0}^{n-1} \theta^i A) > 1 - \epsilon$.

C.4 Discrete Parameter Martingales

C.4.1
Let $(\Omega, \mathscr{E}, \mathbb{P})$ be a probability space and let $(\mathscr{F}_n, n \geq 0)$ be an increasing sequence of sub-σ-fields of $\mathscr{E}$. The sequence $\mathbf{F} = (\mathscr{F}_n, n \geq 0)$ is called a *filtration*. Any sequence of random variables $\{Z_n, n \geq 0\}$ satisfying

(i) Z_n is $\mathscr{F}_n$-measurable $\quad n \in \mathbf{N}$
(ii) $\mathbb{E}|Z_n| < \infty \quad n \in \mathbf{N}$
(iii) $\mathbb{E}(Z_n|\mathscr{F}_m) = Z_m \quad \mathbb{P}-\text{a.s.} \quad m < n, m, n \in \mathbf{N},$

is said to be a $(\mathscr{F}_n)$-*martingale*. Equivalently we say $(Z_n, \mathscr{F}_n; n \geq 0)$ is a martingale.

If (i) and (ii) are retained and (iii) is replaced by the inequality

$$\mathbb{E}(Z_n|\mathscr{F}_m) \geq Z_m \qquad \mathbb{P}-\text{a.s.}$$

(resp. $\mathbb{E}(Z_n|\mathscr{F}_m) \leq Z_m$ $\mathbb{P}$–a.s.), then $(Z_n, \mathscr{F}_n; n \geq 0)$ is called a *submartingale* (resp., a *supermartingale*).

C.4.2
Let $(Z_n, \mathscr{F}_n; n \geq 0)$ be a martingale (or submartingale, or supermartingale). If

$$\sup_n \mathbb{E}|Z_n| < \infty,$$

then (*Doob's martingale convergence theorem*) as $n \to \infty, Z_n$ converges $\mathbb{P}$-a.s. to a limit Z_∞ with $\mathbb{E}|Z_\infty| < \infty$.

C.4.3
Suppose that $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$. Let $(M_n, \mathscr{F}_n; n \geq 0)$ be a martingale such that

$$\sup_n \|M_n\|_p < \infty.$$

Then (*Doob's inequality*)

$$\| \sup_n |M_n| \|_p \leq q \sup_n \|M_n\|_p.$$

Another useful inequality for martingales is the *Burkholder inequality*: if $1 \leq p < \infty$, there exists an absolute constant $C(p)$ such that

$$\| \sup_n |M_n - M_0| \|_p \leq C(p) \left\| \left(\sum_{n=1}^{\infty} (M_n - M_{n-1})^2 \right)^{1/2} \right\|_p.$$

C.4.4
An $(\mathscr{F}_n)$-*stopping time* T is a map $T : \Omega \to \overline{\mathbf{N}} = \mathbf{N} \cup \{+\infty\}$ with

$$\{\omega : T(\omega) \leq n\} \in \mathscr{F}_n, \qquad n \in \mathbf{N}.$$

If $(Z_n, \mathscr{F}_n; n \geq 0)$ is a nonnegative supermartingale, we put

$$Z_T = \sum_{n=0}^{\infty} Z_n 1_{\{T=n\}}$$

and we have

$$\mathbb{E}(Z_0) \geq \mathbb{E}(Z_T) \geq 0.$$

C.5 Continuous Martingales

C.5.1
We suppose now that the index set is $\mathbb{R}_+$ (or a finite interval $[0, T]$). We shall only deal with processes that have right-continuous paths, that is $t \rightsquigarrow Z_t(\omega)$ is *right-continuous* from $\mathbb{R}_+$ to $\mathbb{R}^d$ for every ω in Ω. Two processes with right-continuous paths that are modifications of each other are in fact indistinguishable and will be identified. Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space endowed with a filtration $(\mathscr{F}_t, t \geq 0)$, that is, an increasing family of sub-σ-fields of $\mathscr{A}$. We define, for $t \geq 0$,

$$\mathscr{F}_{t+} = \bigcap_{s>t} \mathscr{F}_s.$$

A filtration $(\mathscr{F}_t)$ is *right-continuous* if $\mathscr{F}_t = \mathscr{F}_{t+}$ for any $t \geq 0$. A d-dimensional random process $Z = (Z_t, t \geq 0)$ is called *adapted* if Z_t is $\mathscr{F}_t$-measurable for every $t \geq 0$, .

C.5.2
A real-valued process Z carried by $(\Omega, \mathscr{A}, \mathbb{P})$ is called an $(\mathscr{F}_t)$-*martingale* if

(i) Z is an $(\mathscr{F}_t)$-adapted random process;
(ii) $\mathbb{E}|Z_t| < \infty, \qquad t \geq 0;$
(iii) $\mathbb{E}(Z_t | \mathscr{F}_s) = Z_s \qquad \mathbb{P}$-a.s., $0 \leq s < t < \infty.$

GCUL

To get an $(\mathcal{F}_t)$-*submartingale* (resp., $(\mathcal{F}_t)$-*supermartingale*) we have to replace the equality in (iii) by the inequality $\geq$ (resp., $\leq$). Recall that the paths are assumed to be right-continuous. This right-continuity allows us to transfer some theorems directly from discrete parameter martingales to continuous parameter ones. This is the case for *Doob's convergence theorem*: If $\sup_t \mathbb{E}|Z_t| < \infty$ and Z is a martingale (or submartingale, or supermartingale), then Z_t converges a.s. as $t \to \infty$ to Z_∞, and $\mathbb{E}|Z_\infty| < \infty$. This is also the case for *Doob's inequality*: If $1 < p < \infty$, $p^{-1} + q^{-1} = 1$ and M is a martingale such that $\sup_t \|M_t\|_p < \infty$, then

$$\| \sup_t |M_t| \|_p \leq q \sup_t \|M_t\|_p.$$

C.5.3

A real-valued process A is *increasing* if it is $(\mathcal{F}_t)$-adapted, has $A_0 = 0$ a.s., and for $\mathbb{P}$-almost every ω in Ω, the path $t \rightsquigarrow A_t(\omega)$ is nondecreasing. With any increasing process A we can associate a random variable $A_\infty : \Omega \to \mathbb{R}_+ \cup \{+\infty\}$ defined by

$$A_\infty = \lim_{t \to \infty} A_t \qquad \mathbb{P} - \text{a.s.}$$

A finite variation process is the difference of two increasing processes.

C.5.4

Let M be a martingale with continuous paths such that for any t in $\mathbb{R}_+$, $\mathbb{E}|M_t|^2 < \infty$. We say that M is a *continuous square integrable martingale*. We define the *quadratic variation process* $\langle M \rangle_t$ to be the unique continuous increasing process A that makes $M_t^2 - A_t$ a continuous martingale. Its existence is ensured by the *Doob-Meyer decomposition of submartingales*. If M and N are both continuous square integrable martingales, the *quadratic covariation process* of M and N is

$$\langle M, N \rangle = \tfrac{1}{4}(\langle M + N \rangle - \langle M - N \rangle).$$

It is easily seen that $MN - \langle M, N \rangle$ is a continuous martingale. We can now state the continuous parameter *Burkholder inequality*: If $1 \leq p < \infty$, there exists an absolute constant $c(p)$ such that for any continuous square integrable martingale M,

$$\| \sup_t |M_t - M_0| \|_p \leq c(p) \, \| (\langle M \rangle_\infty)^{1/2} \|_p .$$

C.5.5

An $(\mathcal{F}_t)$-*stopping time* is a map $T : \Omega \to [0, \infty]$ such that

$$\{\omega : T(\omega) \leq t\} \in \mathcal{F}_t \qquad t \geq 0.$$

Let X be a d-dimensional continuous adapted process and let F be a closed set in $\mathbb{R}^d$. If we set

$$D_F(\omega) = \inf\{t \geq 0 : X_t(\omega) \in F\}$$

with the usual convention that $\inf \varnothing = +\infty$, then D_F is an $(\mathcal{F}_t)$-stopping time. D_F is called the *debut* or the *entrance time* of F. If X is right-continuous adapted and G is an open set in $\mathbb{R}^d$, the time

$$T_G(\omega) = \inf \{t > 0 : X_t(\omega) \in G\}$$

is a stopping time with respect to the filtration $(\mathcal{F}_{t+})$. It is called the *hitting time* of G. Every constant $c \in [0, \infty]$ is a stopping time. If S and T are stopping times, so is $S \wedge T$.

C.5.6

A d-dimensional process $X = (X_t; t \geq 0)$ carried by $(\Omega, \mathcal{A}, \mathbb{P}, (\mathcal{F}_t))$ is called *progressively measurable* if, for every $t \geq 0$, the map

$$(s, \omega) \rightsquigarrow X_s(\omega)$$

from $[0, t] \times \Omega$ to $\mathbb{R}^d$ is $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$-measurable. A right-continuous $(\mathcal{F}_t)$-adapted process X is progressively measurable. Let T be an $(\mathcal{F}_t)$-stopping time. We define the σ-field $\mathcal{F}_T$ by

$$\mathcal{F}_T = \{A \in \mathcal{A} : A \cap \{T \leq t\} \in \mathcal{F}_t \text{ for all } t \geq 0\}.$$

If $X = (X_t; t \geq 0)$ is progressively measurable and X_∞ is any random variable, then

$$X_T(\omega) = X_{T(\omega)}(\omega)1_{\{T<\infty\}}(\omega) + X_\infty(\omega)1_{\{T=\infty\}}(\omega)$$

is an $\mathscr{F}_T$-measurable random variable.

C.5.7
Let M be a (right-continuous) $(\mathscr{F}_t)$-martingale. If S and T are bounded $(\mathscr{F}_t)$-stopping times with $S \leq T$, then (*optional sampling theorem for martingales*)

$$\mathbb{E}[M_T|\mathscr{F}_S] = M_S \qquad \mathbb{P}-\text{a.s.}$$

C.5.8
A real-valued (right-continuous) adapted process X is an $(\mathscr{F}_t)$-*local martingale* if there exist $(\mathscr{F}_t)$-stopping times $T_1 \leq T_2 \leq \cdots$ with $T_n \to \infty \mathbb{P}$-a.s. as $n \to \infty$ such that for every $n \geq 1$ the *stopped process*

$$X_t^{T_n} \equiv X_{t\wedge T_n}$$

is an $(\mathscr{F}_t)$-martingale. If X is a continuous local martingale, there exists a unique continuous increasing process A_t such that $X_t^2 - X_0^2 - A_t$ is a continuous local martingale. It is called the *quadratic variation process* of X and denoted by $\langle X\rangle_t$. This definition is consistent with the one that we have given for continuous square integrable martingales. If X and Y are both continuous local martingales, we still define

$$\langle X, Y\rangle = \tfrac{1}{4}(\langle X+Y\rangle - \langle X-Y\rangle).$$

C.5.9
For $t \in \mathbb{R}_+^*$, a *partition* π_t of $[0,t]$ is a finite ordered subset $\pi_t = \{t_0, \ldots, t_k\}$ of $[0,t]$ such that $0 = t_0 < t_1 < \cdots < t_k = t$. The *mesh* of π_t will be denoted by

$$\delta(\pi_t) = \max\{|t_{j+1} - t_j|; j = 0, \ldots, k-1\}.$$

Suppose M is a continuous local martingale and let $\{\pi_t^n, n \geq 0\}$ be a sequence of partitions of $[0,t]$ such that $\lim_{n\to\infty} \delta(\pi_t^n) = 0$. Then

$$\sum_j (M_{t_{j+1}^n} - M_{t_j^n})^2$$

converges in probability to $\langle M \rangle_t$ as $n \to \infty$. Consequently a continuous local martingale has a.s. infinite variation in any interval $[s,t]$ with $\langle M \rangle_t > \langle M \rangle_s$ a.s.

C.6 Stochastic Integral

C.6.1

If the σ-field $\mathscr{A}$ is $\mathbb{P}$-complete and each $\mathscr{F}_t$ contains all $\mathbb{P}$-null sets of $\mathscr{A}$, then the filtration $(\mathscr{F}_t)$ is said to be $\mathbb{P}$-complete. If M is a continuous local martingale, $L^2_{loc}(M)$ is the space of progressively measurable processes H such that, for every $t \geq 0$,

$$\mathbb{P}\left(\int_0^t H_s^2 d\langle M \rangle_s < \infty \right) = 1$$

If $(\mathscr{F}_t)$ is $\mathbb{P}$-complete and $H \in L^2_{loc}(M)$, there exists a unique continuous local martingale denoted $(H.M)_t$ or $\int_0^t H_s dM_s$ such that for any continuous local martingale N,

$$\langle H.M, N \rangle = H.\langle M, N \rangle$$

where the right-hand side denotes the Stieltjes pathwise integral of the process H with respect to the finite variation process $\langle M, N \rangle$. The process $H.M$ is called the *stochastic integral* of H with respect to M.

C.6.2

A real-valued random process X is a *continuous semi-martingale* if it can be written $X = M + A$, where M is a continuous local martingale and A is a continuous finite variation process; this decomposition is unique. The *quadratic variation process* of X is

$$\langle X \rangle = \langle M \rangle$$

and the *quadratic covariation process* of two continuous semi-martingales $X = M + A$ and $Y = N + B$ is

$$\langle X, Y \rangle = \langle M, N \rangle.$$

C.6.3
A progressively measurable process H is *locally bounded* if there exists a sequence of $(\mathscr{F}_t)$-stopping times $T_1 \leq T_2 \leq \cdots$ with $T_n \to \infty \mathbb{P}$-a.s. as $n \to \infty$ and constants C_n such that the stopped processes H^{T_n} satisfy

$$|H^{T_n}| \leq C_n, \qquad n \geq 1.$$

All continuous adapted processes H are locally bounded, and locally bounded processes are in $L^2_{loc}(M)$ for any continuous local martingale M. Suppose that $(\mathscr{F}_t)$ is $\mathbb{P}$-complete, and let H be a locally bounded process and $X = M + A$ be a continuous semi-martingale; then the *stochastic integral* of H with respect to X is the continuous semi-martingale

$$H.X = H.M + H.A,$$

where $H.M$ is the stochastic integral in C.6.1 and $H.A$ is a Stieltjes pathwise integral. We will also write $(H.X)_t = \int_0^t H_s dX_s$.

C.6.4
If X and Y are two continuous semi-martingales, then (*integration by parts formula*)

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + \langle X, Y \rangle_t.$$

This equation is a particular case of *Ito's formula* (*chain rule for stochastic integrals*): Let $F \in \mathscr{C}^2(\mathbb{R}^d; \mathbb{R})$ and let $X = (X^1, \ldots, X^d)$ be a continuous d-dimensional semi-martingale; then $F(X)$ is a continuous semi-martingale and

$$F(X_t) = F(X_0) + \sum_i \int_0^t \frac{\partial F}{\partial x_i}(X_s)\, dX_s^i + \frac{1}{2} \sum_{i,j} \int_0^t \frac{\partial^2 F}{\partial x_i \partial x_j}(X_s)\, d\langle X^i, X^j \rangle_s.$$

C.6.5
Let X and Y be continuous semi-martingales. The *Stratonovich stochastic integral* is defined by

$$\int_0^t X_s \circ dY_s = \int_0^t X_s dY_s + \tfrac{1}{2}\langle X, Y\rangle_t.$$

If $F \in \mathscr{C}^3(\mathbb{R}^d; \mathbb{R})$ and $X = (X^1, \ldots, X^d)$ is a d-dimensional semi-martingale, then we have the usual chain rule:

$$F(X_t) = F(X_0) + \int_0^t \frac{\partial F}{\partial x_i}(X_s) \circ dX_s^i.$$

D WIENER-LEVY CALCULUS

D.1 Brownian Motion

D.1.1
Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space. A *d-dimensional Brownian motion* $(B_t, t \geq 0)$ is a d-dimensional random process that has the following properties:

(i) (independent increments) if $0 \leq t_0 < t_1 \cdots < t_n$, then $B_{t_0}, B_{t_1} - B_{t_0}, \cdots, B_{t_n} - B_{t_{n-1}}$ are independent;
(ii) if $s, t \geq 0$, then $B_{t+s} - B_s$ has distribution $N_d(0, tI)$, where I is the identity matrix;
(iii) with probability 1, $t \rightsquigarrow B_t$ is continuous.

If moreover $B_0 = 0$ $\mathbb{P}$-a.s., then the Brownian motion is said to be *standard*. If $B_0 = x$ $\mathbb{P}$-a.s., where $x \in \mathbb{R}^d$, then it is said to *start at x*.

D.1.2
If B is a standard d-dimensional Brownian motion, then the following properties hold:

(i) (Homogeneity). For any $s > 0$, the process $(B_{t+s} - B_s, t \geq 0)$ is a standard Brownian motion independent of $\sigma(B_u, u \leq s)$.
(ii) (Symmetry). The process $(-B_t, t \geq 0)$ is a standard Brownian motion.
(iii) (Scaling). For every $c > 0$, the process $(cB_{t/c^2}, t \geq 0)$ is a standard Brownian motion.

D.1.3
In some cases, the interesting filtration will be larger than $(\sigma(B_u, u \le t), t \ge 0)$, and we shall adapt our definition. A d-dimensional $(\mathscr{F}_t)$-Brownian motion is an $\mathbb{R}^d$-valued random process that has the following properties:

(i) $(B_t, t \ge 0)$ is an $(\mathscr{F}_t)$-adapted process;
(ii) if $s, t \ge 0$, then $B_{t+s} - B_s$ is independent of $\mathscr{F}_s$ and has distribution $N_d(0, tI)$;
(iii) with probability $1, t \rightsquigarrow B_t$ is continuous.

In the above definition, we may replace (ii) by the equivalent condition:

$$\mathbb{E}[\exp\{i\langle\xi, B_{t+s} - B_s\rangle | \mathscr{F}_s\}] = \exp\{-t|\xi|^2/2\} \qquad \mathbb{P} - \text{a.s.}$$

for all $\xi \in \mathbb{R}^d$ and $s, t \ge 0$.

D.1.4
A *Gaussian process* with index set T is a collection of real random variables $(X_t; t \in T)$ whose finite-dimensional distributions are all Gaussian. A real Brownian motion starting at x is a Gaussian process with moments

$$\mathbb{E}(B_t) = x \qquad t \ge 0$$
$$\text{Cov}(B_s, B_t) = s \wedge t \qquad s, t \ge 0.$$

If φ is a bounded real Borel function and $t > 0$, we have

$$\mathbb{E}(\varphi(B_t)) = \int \varphi(y) p(t; x, y)\, dy$$

where

$$p(t; x, y) = (2\pi t)^{-1/2} \exp\{-(x - y)^2/2t\}$$

is the *transition-density function* of the Brownian motion.

D.1.5
Another important Gaussian process is the centered Gaussian process defined on $T = [0, 1]$ with covariance

$$\text{Cov}(X_s, X_t) = s(1 - t) \qquad \text{if} \quad 0 \le s \le t \le 1.$$

The easiest way to get such a process is to set

$$X_t = B_t - tB_1, \qquad 0 \le t \le 1,$$

where B is a standard one-dimensional Brownian motion. This immediately gives us a continuous version of the process, which is called a *Brownian bridge*. More generally, one may consider, for x, y in $\mathbb{R}$,

$$X_t^{x,y} = x + B_t + t(y - x - B_1).$$

This process is called a *pinned Brownian motion*. It is easy to see that if $\varphi_1, \ldots, \varphi_n, \psi$ are bounded Borel functions on $\mathbb{R}$, and $0 < t_1 < \cdots < t_n < 1$, then

$$\begin{aligned}\mathbb{E}[\varphi_1(x + B_{t_1}) \cdots \varphi_n(x + B_{t_n})\psi(x + B_1)] \\ = \int \mathbb{E}[\varphi_1(X_{t_1}^{x,y}) \cdots \varphi_n(X_{t_n}^{x,y})]\psi(y)p(1; x, y)\, dy.\end{aligned}$$

In some sense, the law of $X^{x,y}$ coincides with the law of the Brownian motion starting at x and conditioned on being at y at time 1.

D.1.6

A third useful Gaussian process is obtained in the following way. Let B be a standard one-dimensional Brownian motion. The process with index set $\mathbb{R}$ defined by

$$V_t = e^{-t} B_{e^{2t}} \qquad t \in \mathbb{R},$$

is called an *Ornstein-Uhlenbeck process*. It is a Gaussian process with zero mean and covariance

$$\mathbb{E}(V_t V_s) = e^{-|t-s|}, \qquad s, t \in \mathbb{R}.$$

The probability distribution of V_t is $N(0, 1)$ for any $t \in \mathbb{R}$ and for any bounded Borel functions φ and ψ on $\mathbb{R}$, we have, for $s \neq t$,

$$\mathbb{E}[\varphi(V_s)\psi(V_t)] = \iint \varphi(x)\psi(y)p(1; 0, x)q(t - s; x, y)\, dx\, dy$$

where, for $u \neq 0$, $q(u; x, .)$ is the density function of the probability distribution $N(e^{-|u|}x, 1 - e^{-2|u|})$.

D.1.7
Let B be a standard d-dimensional $(\mathscr{F}_t)$-Brownian motion. The following processes are $(\mathscr{F}_t)$-martingales:

(i) $B_t^i \qquad i \in \{1, \ldots, d\}$
(ii) $B_t^i B_t^j - \delta_i^j t \qquad i,j \in \{1, \ldots, d\}$
(iii) $\exp\{\langle \theta, B_t \rangle - \frac{1}{2}|\theta|^2 t\} \qquad \theta \in \mathbb{R}^d.$

A nontrivial result is that a continuous process B with $\mathbb{P}(B_0 = 0) = 1$ is a standard d-dimensional $(\mathscr{F}_t)$-Brownian motion if and only if (i) and (ii) hold. Property (iii) can be extended to the complex case:

$$\exp\{i\langle u, B_t \rangle + \tfrac{1}{2}|u|^2 t\}$$

is a complex $(\mathscr{F}_t)$-martingale for every u in $\mathbb{R}^d$.

D.1.8
Properties (i) and (ii) above have two important consequences:

- $\langle B^i \rangle_t = t$ for $i \in \{1, \ldots, d\}$ and so almost every Brownian path has unbounded variation on every interval;
- if $(\mathscr{F}_t)$ is $\mathbb{P}$-complete, $H.B \equiv \sum_i H^i.B^i$ may be defined for any d-dimensional progressively measurable H such that

$$\mathbb{P}\left(\int_0^t |H_s|^2 \, ds < \infty \right) = 1 \qquad \forall t \geq 0.$$

D.1.9
Suppose that the filtration $(\mathscr{F}_t)$ is $\mathbb{P}$-complete, and let B be a standard d-dimensional $(\mathscr{F}_t)$-Brownian motion. Let $T > 0$ and let H be a d-dimensional progressively measurable process satisfying

$$\mathbb{P}\left[\int_0^T |H_s|^2 \, ds < \infty \right] = 1.$$

For $0 \le t \le T$, set

$$Z_t(H) = \exp\left\{\sum_{i=1}^{d}\int_0^t H_s^i dB_s^i - \tfrac{1}{2}\int_0^t |H_s|^2\, ds\right\}$$

As a result of Ito's formula, the process $(Z_t(H); 0 \le t \le T)$ is a one-dimensional continuous local martingale with $Z_0(H) = 1$. It is a martingale if and only if

$$\mathbb{E}[Z_T(H)] = 1.$$

In this case, we define a probability measure $\tilde{P}_T$ on $(\Omega, \mathscr{F}_T)$ by

$$\tilde{P}_T(A) = \mathbb{E}[1_A Z_T(H)], \qquad A \in \mathscr{F}_T$$

and the process $\tilde{B} = (\tilde{B}^1, \ldots, \tilde{B}^d)$ defined by

$$\tilde{B}_t^i = B_t^i - \int_0^t H_s^i ds \qquad 1 \le i \le d, \quad 0 \le t \le T$$

is a standard d-dimensional Brownian motion carried by $(\Omega, \mathscr{F}_T, \tilde{P}_T)$ (*Cameron–Martin–Girsanov theorem*). A sufficient condition for $Z(H)$ to be a martingale is Novikov's condition:

$$\mathbb{E}[\exp \tfrac{1}{2}\int_0^T |H_s|^2 ds] < \infty.$$

D.1.10

In many electronic or mechanical systems, the sensitivity of the devices is limited by continual fluctuations, which we call noise. Thus it would be interesting to represent it by a centered Gaussian process with covariance $1_{\{s=t\}}$. For such a process, any two random variables X_s, X_t with $s \ne t$ would be independent. Now, there is no such measurable map $(t, \omega) \rightsquigarrow X_t(\omega)$ since, by Fubini's theorem,

$$Y_t = \int_0^t X_s \, ds$$

would be a centered Gaussian random variable with $E(Y_t^2) = 0$, and hence $X_t = 0\,\lambda_1 \otimes \mathbb{P}$-a.e.; on the other hand,

$$\mathbb{E}\left[\int_0^t X_s^2 ds\right] = t,$$

which gives a contradiction. Such a process X may, however, be treated as a functional that acts on functions f in $L^2_{loc}(\mathbb{R}_+)$

$$\int_0^t f(s) X_s \, ds = \int_0^t f(s) dB_s,$$

where B is a standard Brownian motion. The process X is called the *white noise process*. It is the derivative of B in the sense of Schwartz distributions.

D.1.11

Any continuous local martingale whose quadratic variation process tends to infinity can be time changed to a Brownian motion. We state the multivariate version of this property, which is *Knight's theorem*. Suppose that $(\mathcal{F}_t)$ is a right-continuous filtration and let $M = (M^1, \ldots, M^d)$ be a d-dimensional continuous local martingale with

$$\begin{aligned} M_0^i &= 0 \quad \mathbb{P}\text{-a.s.} \qquad i = 1, \ldots, d \\ \langle M^i \rangle_\infty &= \infty \quad \mathbb{P}\text{-a.s.} \qquad i = 1, \ldots, d \\ \langle M^i, M^j \rangle_t &= 0 \quad \mathbb{P}\text{-a.s.} \qquad 1 \le i \ne j \le d, \quad 0 \le t < \infty. \end{aligned}$$

Let

$$T^i(s) = \inf\{t \ge 0 : \langle M^i \rangle_t > s\} \qquad 0 \le s < \infty, \quad 1 \le i \le d.$$

Then the process $B = (B^1, \ldots, B^d)$ defined by

$$B_s^i = M_{T^i(s)}^i \qquad 0 \le s < \infty; \quad 1 \le i \le d$$

is a standard d-dimensional Brownian motion.

D.2 Wiener Space

D.2.1
Let $\mathscr{C}(\mathbb{R}_+; \mathbb{R}^d)$ be the set of all continuous functions $w : \mathbb{R}_+ \to \mathbb{R}^d$. We define a metric on $\mathscr{C}(\mathbb{R}_+; \mathbb{R}^d)$ by

$$\rho(w_1, w_2) = \sum_{n=1}^{\infty} 2^{-n}[(\sup_{0 \le t \le n} |w_1(t) - w_2(t)|) \wedge 1] \qquad w_1, w_2 \in \mathscr{C}(\mathbb{R}_+; \mathbb{R}^d).$$

This space is complete and separable under the metric ρ. Let $\pi = \{\pi_t; t \ge 0\}$ be the collection of canonical projections

$$\pi_t(w) = w(t), \qquad t \in \mathbb{R}_+, \quad w \in \mathscr{C}(\mathbb{R}_+; \mathbb{R}^d).$$

If $\mathscr{B}$ denotes the Borel σ-field on $\mathscr{C}(\mathbb{R}_+; \mathbb{R}^d)$, then $\sigma(\pi) = \mathscr{B}$, and every probability measure on $(\mathscr{C}(\mathbb{R}_+; \mathbb{R}^d), \mathscr{B})$ is uniquely determined by its values on the sets of form:

$$\{w : (w(t_1), \ldots, w(t_n)) \in E\},$$

where $0 \le t_1 < \cdots < t_n$ and $E \in \mathscr{B}(\mathbb{R}^{nd})$.

D.2.2
Let B be a d-dimensional Brownian motion starting at $x \in \mathbb{R}^d$. As B is $\mathbb{P}$-a.s. a mapping from Ω into $\mathscr{C}(\mathbb{R}_+; \mathbb{R}^d)$, we may consider its law $\mathbb{P}_x = B_* \mathbb{P}$, which does not depend on the particular Brownian motion B. We call $\mathbb{P}_x$ the *Wiener measure started at* x or, if $x = 0$, shortly the *Wiener measure*. The probability space $(\mathscr{C}(\mathbb{R}_+; \mathbb{R}^d), \mathscr{B}, \mathbb{P}_0)$ is called the *Wiener space*.

D.2.3
In some cases it will be more convenient to restrict the parameter set to $[0, 1]$ and deal with the sup norm and the Borel σ-field $\mathscr{B}$, or even to restrict to the closed subset $\mathscr{C}_0([0, 1]; \mathbb{R}^d)$ of continuous functions that have $w(0) = 0$. The *Cameron–Martin space* is the Hilbert space of continuous functions

$$f : [0, 1] \to \mathbb{R}^d$$

which can be written

$$f(t) = \int_0^t \dot{f}(s)\, ds, \qquad 0 \le t \le 1,$$

for some $\dot{f}$ in $L^2([0, 1]; \mathbb{R}^d)$. It has symmetric form

$$\langle f, g \rangle_H = \int_0^1 \langle \dot{f}(s), \dot{g}(s) \rangle ds.$$

The Cameron–Martin space is a dense subset of $\mathscr{C}_0([0, 1]; \mathbb{R}^d)$.

D.2.4

Now suppose that we are working in the canonical setting : the probability space $(\Omega, \mathscr{A})$ is $(\mathscr{C}(\mathbb{R}_+; \mathbb{R}^d), \mathscr{B})$ and for any probability measure $\mathbb{P}_x$ on $(\Omega, \mathscr{A})$, the process of canonical mappings $(B_t, t \ge 0)$ is a Brownian motion starting at x. For $s \ge 0$ we define the *shift transformation* θ_s by setting

$$(\theta_s \omega)(t) = \omega(s + t), \qquad t \ge 0.$$

The canonical filtration on Ω is $\{\mathscr{F}_t = \sigma(B_s; s \le t); t \ge 0\}$ and we have the following *Markov property*: If $s \ge 0$ and Y is a bounded random variable, then for all $x \in \mathbb{R}^d$,

$$\mathbb{E}_x[Y \circ \theta_s | \mathscr{F}_s] = \mathbb{E}_{B_s} Y \qquad \mathbb{P}_x - \text{a.s.},$$

where the right-hand side is the function $\varphi(x) = \mathbb{E}_x Y$ evaluated at $x = B_s$. The following consequence is particularly useful: Let F be a closed set in $\mathbb{R}^d$, let Z be a bounded $\mathscr{F}_s$-measurable random variable, and let g be a bounded Borel function on $\mathbb{R}^d$. If D_F denotes the debut of F, then

$$D_F = s + D_F \circ \theta_s \qquad \text{on} \quad \{D_F > s\}$$

and

$$\mathbb{E}_x[Z g(B_{D_F}) 1_{\{D_F > s\}}] = \mathbb{E}_x[Z 1_{\{D_F > s\}} \mathbb{E}_{B_s}[g(B_{D_F})]].$$

CHAPTER 2

Computation of Expectations in Finite Dimension

In this chapter we lay out the mathematical framework for studying Monte Carlo and quasi-Monte Carlo methods, which are used to calculate the expectation of random variables defined on a finite dimensional space. We also consider the numerical computation of conditional expectations.

A MATHEMATICAL FRAMEWORK OF SIMULATION

> There is probably no mathematical subject which shares with probability the features that on the one hand many of its most elementary theorems are based on rather deep mathematics, and that on the other hand many of its most advanced theorems are known and understood by many (statisticians and others) without the mathematical background to understand their proofs. It follows that any rigorous treatment of probability is threatened with the charge that it is absurdly overloaded with extraneous mathematics. [J. L. Doob, 1952]

A.1 Narrow Convergence

We begin by summarizing some facts about narrow convergence.

Let E be a Polish space (a complete separable metric space), and let $\mathscr{E}$ be the σ-field generated by the open sets in E. We call $\mathscr{E}$ the Borel σ-field on E.

A.1.1 Examples

The sets $[0,1]^s$, $\{0,1\}^s$, for $s \in \mathbf{N}^*$, $[0,1]^{\mathbf{N}}$, $\{0,1\}^{\mathbf{N}}$, all endowed with their product topologies, are Polish spaces and compact Hausdorff spaces. $\mathbb{R}^s$ and $\mathbb{R}_+^s$ are locally compact Hausdorff spaces. For $d \in \mathbf{N}^*$, $\mathscr{C}(\mathbb{R}_+, \mathbb{R}^d)$ is a Polish space that is an infinite-dimensional Fréchet space (complete metrizable locally convex space).

Definition A.1.2 On the set $\mathcal{M}_+^1(E, \mathcal{E})$ of probability measures on $(E, \mathcal{E})$, the *narrow topology* is the coarsest topology for which the mappings: $\mu \rightsquigarrow \mu(f)$ are continuous for all f in the set $\mathcal{C}_b(E)$ of bounded continuous functions on $(E, \mathcal{E})$.

A.1.3 Exercise

Let $(\mu_n)_{n\geq 1}$ be a sequence in $\mathcal{M}_+^1(E, \mathcal{E})$. Suppose that the sequence $(\mu_n(f))_{n\geq 1}$ is convergent for every bounded function f. Show that there is some μ in $\mathcal{M}_+^1(E, \mathcal{E})$ such that $\mu_n(f)$ converges to $\mu(f)$ for every continuous bounded f.

A.1.4 Exercise

Show that there exists a countable subset $\mathcal{H}$ of $\mathcal{C}_b(E)$ such that, if μ_n, $\mu \in \mathcal{M}_+^1(E, \mathcal{E})$ and $\mu_n(f) \to \mu(f)$ for all f in $\mathcal{H}$, then $\mu_n \to \mu$ narrowly. We call it a *convergence-determining set.*

Definition A.1.5 A subset $\mathcal{A}$ of $\mathcal{M}_+^1(E, \mathcal{E})$ is called *tight* if, for every $\epsilon > 0$, there exists a compact subset K of E such that $\mu(E\backslash K) < \epsilon$ for every μ in $\mathcal{A}$.

Theorem A.1.6 (Prokhorov) Let $\mathcal{A}$ be a subset of $\mathcal{M}_+^1(E, \mathcal{E})$. The following statements are equivalent:

(i) every sequence of elements of $\mathcal{A}$ contains a convergent subsequence,

(ii) $\mathcal{A}$ is tight,

(iii) $\mathcal{A}$ is relatively compact.

A.2 Narrow Convergence and μ-Riemann Integrable Functions

Definition A.2.1 Let μ be a probability measure on $(E, \mathcal{E})$. A real bounded Borel function f on E is called μ-*Riemann integrable* if, for every $\epsilon > 0$, there exist continuous bounded functions u and v on E such that $u \leq f \leq v$ and $\int (v - u) d\mu \leq \epsilon$. If f is $\mathcal{R}^d$-valued, then f is said to be μ-Riemann integrable if its components are real μ-Riemann integrable functions.

A.2.2 Exercise

Let f be a Borel bounded function from $(E, \mathcal{E})$ into $\mathcal{R}^d$. Prove that f is μ-Riemann integrable iff for every $\epsilon > 0$, there exists g in $\mathcal{C}_b(E, \mathcal{R}^d)$ and h in $\mathcal{C}_b(E)$ such that

$$|f - g| \leq h \qquad \text{and} \qquad \int h d\mu \leq \epsilon.$$

Theorem A.2.3 Let $(\mu_n)_{n\geq 1}$ be a sequence of probability measures on $(E, \mathscr{E})$, which converges narrowly to a limit μ. Then $\int f d\mu_n \to \int f d\mu$ for every μ-Riemann integrable function f.

Proof We may assume that f is a real function. Given $\epsilon > 0$, let $u, v \in \mathscr{C}_b(E)$ satisfy $u \leq f \leq v$ and $\int (v - u) d\mu \leq \epsilon$. Then

$$\int u d\mu_n \leq \int f d\mu_n \leq \int v d\mu_n.$$

The left- and right-hand terms converge to $\int u d\mu$ and $\int v d\mu$ respectively, so

$$\overline{\lim} \int f d\mu_n - \underline{\lim} \int f d\mu_n \leq \epsilon.$$

Since ϵ is arbitrary, we deduce that $\int f d\mu_n$ converges; and from the inequalities we see that the limit is $\int f d\mu$. □

Theorem A.2.4 Let f be a real bounded Borel function on $(E, \mathscr{E})$. Then f is μ-Riemann integrable iff the set of its points of discontinuity is μ-negligible.

Proof Let us introduce the lower and upper regularized functions:

$$\check{f}(x) = \overline{\lim_{y \to x}} f(y) \qquad \text{and} \qquad \hat{f}(x) = \underline{\lim_{y \to x}} f(y).$$

Then $\hat{f} \leq f \leq \check{f}$, and f is continuous iff $\hat{f} = \check{f}$. We note that $\check{f}$ is upper semi-continuous (usc) bounded and $\hat{f}$ is lower semi-continuous (lsc) bounded.

(a) Suppose that the set of points of discontinuity of f is μ-negligible. Then

$$\int \hat{f} d\mu = \int f d\mu = \int \check{f} d\mu.$$

Since $\hat{f}$ is lsc bounded and $\check{f}$ is usc bounded, there exist sequences $(\varphi_n), (\psi_n)$ in $\mathscr{C}_b(E)$ such that $\varphi_n \uparrow \hat{f}$ and $\psi_n \downarrow \check{f}$. Therefore,

$$\int \varphi_n d\mu \uparrow \int \hat{f}\, d\mu = \int f\, d\mu$$

and

$$\int \psi_n d\mu \downarrow \int \check{f}\, d\mu = \int f\, d\mu.$$

We deduce that f is μ-Riemann integrable.

(b) Conversely, suppose that f is μ-Riemann integrable. Given $\epsilon > 0$. let u and v be bounded continuous functions such that $u \le f \le v$ and $\int (v-u)\, d\mu \le \epsilon$. From the continuity of u and v we have

$$u \le \hat{f} \le f \le \check{f} \le v,$$

and therefore

$$\int (\check{f} - \hat{f}) d\mu \le \epsilon.$$

Since ϵ is arbitrary, we deduce that $\mu\{x : \hat{f}(x) < \check{f}(x)\} = 0$. □

A.2.5 Examples

(1) Let (X_n) be a sequence of i.i.d. $\mathbb{R}^d$-valued random variables such that $\mathbb{E}|X_1|^2 < \infty$, and let

$$m = \mathbb{E}X_1$$

and

$$D = \mathbb{E}[(X_1 - m)^t (X_1 - m)].$$

Let $S_n = X_1 + \cdots + X_n$ and $\mu = N_d(0, D)$. Then

$$\mathbb{P}\left\{ \frac{S_n - nm}{\sqrt{n}} \in A \right\} \to \mu(A),$$

for every Borel set $A \subseteq \mathbb{R}^d$ with μ-negligible boundary (Lebesgue-negligible if the matrix D is regular).

(2) Let $(X_n)_{n\geq 1}$ be a sequence of $\mathbb{R}^d$-valued random variables. For $x = (x_1, \ldots, x_d)$, let

$$F_{X_n}(x) = \mathbb{P}\left(\bigcap_{i=1}^{d} \{X_n^i \leq x_i\} \right)$$

be the distribution function of X_n. If (X_n) converges in law to X (i.e. $\mathbb{P} \circ X_n^{-1}$ converges narrowly to $\mathbb{P} \circ X^{-1}$), then (F_{X_n}) converges to F_X at each continuity point x of F_X.

A.2.6 Exercise

Show that, conversely, if (F_{X_n}) converges to F_X at each point x in a countable dense set in $\mathbb{R}^d$, then (X_n) converges in law to X.

A.3 Application to Simulation

> There is a well known mathematical device to map a square onto a line segment in such a way that area goes into length. All we need to do is to write our coordinates in the square in the decimal form:
>
> $$\alpha = .\alpha_1\alpha_2 \ldots \alpha_n \ldots ; \beta = .\beta_1\beta_2 \ldots \beta_n \ldots ;$$
>
> and to put
>
> $$\gamma = .\alpha_1\beta_1\alpha_2\beta_2 \ldots \alpha_n\beta_n \ldots ;$$
>
> and we obtain a mapping of this sort which is one to one for almost all points both in the line-segment and in the square. [Norbert Wiener, 1947]

The following result is sometimes called the *fundamental theorem of simulation.*

Theorem A.3.1 Let E be a Polish space with Borel σ-field $\mathscr{E}$. For every probability measure μ on $(E, \mathscr{E})$ and for every s in $\mathbf{N}^*$, there exists a Borel function f on $[0, 1]^s$ such that f is λ_s-a.e. continuous and $\mu = \lambda_s \circ f^{-1}$.

Here λ_s denotes the s-dimensional Lebesgue measure on $[0, 1]^s$. We state a preliminary lemma before proving the theorem.

Lemma A.3.2 There exist Borel functions $h: [0, 1] \to [0, 1]^{\mathbf{N}}$ and $k: [0, 1]^{\mathbf{N}} \to [0, 1]$ such that $h \circ k = Id$ outside a countable set $A \subset [0, 1]^{\mathbf{N}}$, and h is continuous outside a countable set $B \subset [0, 1]$.

Proof of the Lemma We define the mappings

- $h_1: [0, 1] \to \{0, 1\}^{\mathbf{N}}$
- $h_2: \{0, 1\}^{\mathbf{N}} \to (\{0, 1\}^{\mathbf{N}})^{\mathbf{N}}$
- $h_3: (\{0, 1\}^{\mathbf{N}})^{\mathbf{N}} \to [0, 1]^{\mathbf{N}}$

by

- $(h_1(1))_n = 1 \quad \forall n \in \mathbf{N}$
- $(h_1(x))_n = q_n \quad \forall n \in \mathbf{N}$ if $x < 1$, where (q_n) is the canonical binary expansion $x = \sum_{n=0}^{\infty} (q_n/2^n)$, with the restriction card $\{n: q_n = 0\} = +\infty$
- h_2 is induced by any one-to-one mapping from $\mathbf{N}$ onto $\mathbf{N} \times \mathbf{N}$
- $h_3((q_{m,n})_{m,n\geq 0}) = (\sum_{n=0}^{\infty} (q_{m,n}/2^{m+n}))_{m\geq 0}$.

In the same way, we define the mappings

- $k_1: \{0, 1\}^{\mathbf{N}} \to [0, 1]$
- $k_2: (\{0, 1\}^{\mathbf{N}})^{\mathbf{N}} \to \{0, 1\}^{\mathbf{N}}$
- $k_3: [0, 1]^{\mathbf{N}} \to (\{0, 1\}^{\mathbf{N}})^{\mathbf{N}}$

by

- $k_3((x_n)_{n\geq 0}) = (h_1(x_n))_{n\geq 0}$
- $k_2 = h_2^{-1}$
- $k_1((q_n)_{n\geq 0}) = \sum_{n=0}^{\infty} (q_n/2^{n+1})$.

The assertion of the lemma is now easily verified by setting:

- $h = h_3 \circ h_2 \circ h_1$
- $k = k_1 \circ k_2 \circ k_3$
- $A = (k_2 \circ k_3)^{-1}((h_1([0, 1]))^c)$

and taking B to be the set of dyadic rationals in [0, 1]. □

We can now prove Theorem A.3.1.

Proof of the Theorem It is enough to consider the case $s = 1$. We know [cf. Bourbaki, *Topologie Générale*, Chapter 9, paragraph 6] that a Polish space is homeomorphic to a Borel subset of $[0, 1]^{\mathbf{N}}$, so we may assume that

$$(E, \mathscr{E}) = ([0,1]^{\mathbf{N}}, \mathscr{B}([0,1]^{\mathbf{N}})).$$

If h, k, A, and B are as above, let us define a new countable set

$$C = A \cup h(k(A^c) \cap B).$$

This satisfies $k(C^c) \cap B = \varnothing$ and $h \circ k = Id$ outside C. Let μ be a probability measure on $[0,1]^{\mathbf{N}}$; we decompose μ as

$$\mu = 1_{C^c}.\mu + 1_C.\mu.$$

The measure $\nu = 1_{C^c}.\mu$ can be simulated in the following way: for $x \in [0, \mu(C^c)]$, let

$$u(x) = \inf\{y \in [0,1] : \nu \circ k^{-1}([0,y]) \geq x\}.$$

It is well known that, for every Borel set D in $[0,1]$,

$$\lambda_1(u^{-1}(D)) = \nu(k^{-1}(D)),$$

and in particular

$$\lambda_1(u^{-1}(B)) = \nu(k^{-1}(B)) = \mu(C^c \cap k^{-1}(B)) = 0.$$

Since u is non decreasing, $f_1 = h \circ u$ is λ_1-a.e. continuous, and $\lambda_1 \circ f_1^{-1} = \nu$ on C^c. Now $1_C.\mu$ is supported by a countable set and easily simulated by a mapping f_2 from $]\mu(C^c), 1]$ into C that takes constant values on intervals with suitable lengths. The function f, defined to be f_1 on $[0, \mu(C^c)]$ and f_2 on $]\mu(C^c), 1]$, has the required properties, so we are done. □

A.4 Distribution Functions

For $x = (x_1, \ldots, x_s)$ and $y = (y_1, \ldots, y_s)$ in $\mathbb{R}^s$, we write $x \leq y$ if $x = x \wedge y$. We write $[[x, y]]$ for the the s-dimensional interval $\prod_{i=1}^s [x_i, y_i]$ in $\mathbb{R}^s$, while $]]x, y]]$ denotes $\prod_{i=1}^s]x_i, y_i]$. This last notation will also be used when one or several of the x_i's are $-\infty$.

Let μ be a probability measure on $(\mathbb{R}^s, \mathscr{B}(\mathbb{R}^s))$ with distribution function

$$F_\mu(x) = \mu(]] - \infty, x]]), \qquad x \in \mathbb{R}^s.$$

Lemma A.4.1

$$(*) \quad \begin{cases} F_\mu \text{ is nonnegative, nondecreasing in each variable and} \\ \text{right-continuous by upper sector: if } \alpha = (\alpha_1, \ldots, \alpha_s) \text{ and} \\ x = (x_1, \ldots, x_s), \\ \quad \lim_{x_1 \downarrow \alpha_1, \ldots, x_s \downarrow \alpha_s} F_\mu(x) = F_\mu(\alpha), \\ \quad \lim_{x_1 \uparrow \infty, \ldots, x_s \uparrow \infty} F_\mu(x) = 1 \\ \quad \lim_{x_i \downarrow -\infty} F_\mu(x) = 0 \qquad \text{for every} \quad i = 1, \ldots, s. \end{cases}$$

$$(**) \quad \begin{cases} \text{By setting, for } a < b, \\ [F_\mu(x_1, \ldots, x_s)]_{x_k=a}^{x_k=b} = \\ F_\mu(x_1, \ldots, x_{k-1}, b, x_{k+1}, \ldots, x_s) - F_\mu(x_1, \ldots, x_{k-1}, a, x_{k+1}, \ldots, x_s), \\ \text{for } a_1 < b_1, \ldots, a_s < b_s, \text{ we get} \\ [\ldots[[F_\mu(x_1, \ldots, x_s)]_{x_1=a_1}^{x_1=b_1}]_{x_2=a_2}^{x_2=b_2} \ldots]_{x_s=a_s}^{x_s=b_s} \geq 0. \end{cases}$$

The proof is straightforward.

Lemma A.4.2 Conversely, if a function F from $\mathbb{R}^s$ into [0, 1] satisfies (*) and (**), then there exists a probability measure μ on $\mathbb{R}^s$ such that $F = F_\mu$.

Proof A probabilistic proof can be found in the books by L. Breiman [1968], E.J. McShane [1974] or A.N. Shiryayev [1984]. We give a proof that relies on the theory of Schwartz distributions. Let T be the distribution

$$\frac{\partial^s}{\partial x_1 \cdots \partial x_s} F,$$

and let α be a positive element of the space $\mathscr{D}$ of test functions. Consider the convolution

$$(\alpha * F)(x) = \int \alpha(y) F(x - y)\, dy.$$

The function $\alpha * F$ satisfies also (**): if r is an s-dimensional interval, we have

$$\int_r \frac{\partial^s}{\partial x_1 \cdots \partial x_s} \alpha * F \, d\lambda_s \geq 0.$$

Then $\int_r \alpha * T \, d\lambda_s \geq 0$, and T is a positive measure. Suppose in addition that $\int \alpha \, d\lambda_s = 1$ and support $(\alpha) \subset]-\infty, 0[^s$, and define:

$$\alpha_n(x) = n^s \alpha(nx).$$

By (*), $\alpha_n * F \downarrow F$ as $n \to \infty$. Using Fubini's theorem, we get

$$\begin{aligned}
\alpha_n * F(x) &= \int_{]]-\infty, x]]} \alpha_n * T \, d\lambda_s \\
&= \int 1_{]]-\infty, x]]}(z) \int \alpha_n(z-y) T(dy)\, dz \\
&= \int T(dy) \int 1_{]]-\infty, x]]}(u+y) \alpha_n(u)\, du \\
&= (\alpha_n * T(]]-\infty, .]]))(x).
\end{aligned}$$

The nondecreasing function $F_T(x) = T(]]-\infty, x]])$ is a.e. finite, and thus everywhere finite and right-continuous by upper sector.

From $\alpha_n * F = \alpha_n * F_T$, we deduce that $F = F_T$ by letting n tend to infinity. □

Lemma A.4.3 Let $(F_n)_{n \geq 1}$ be a sequence of distribution functions on $[0,1]^s$, and let π be the distribution function $\pi(x) = x_1 \cdots x_s$ of the Lebesgue measure λ_s on $[0,1]^s$. If $F_n \to \pi$ in $L^1([0,1]^s, \lambda_s)$, then $F_n \to \pi$ uniformly.

Proof (a) Let F be the distribution function of a probability measure μ on $[0,1]^s$ and let $a \in]0, \frac{1}{2}[$. If, for some $b > 0$ and some $x_0 \in [a, 1-a]^s$, we have $F(x_0) \leq \pi(x_0) - b$, then

$$\int |F(x) - \pi(x)| dx \geq \int (\pi(x) - (\pi(x_0) - b))^+ 1_{[[0, x_0]]}(x)\, dx.$$

We get an analogous inequality if $F(x_0) \geq \pi(x_0) + b$, and deduce that if $|F(x_0) - \pi(x_0)| \geq b$ for some $b > 0$ and some $x_0 \in [a, 1-a]^s$, then there exists $C(a,b) > 0$ such that

$$\int |F(x) - \pi(x)| dx \geq C(a,b).$$

Therefore, for every $b > 0$, the inequality

$$\int |F(x) - \pi(x)| dx < C(a,b)$$

implies that, for every $x \in [a, 1-a]^s$,

$$\int |F(x) - \pi(x)| < b.$$

(b) Let $A = [a, 1-a]^s$ and assume $|F - \pi| < \epsilon$ on A. If $x_i < a$ for some $i \in \{1, \ldots, s\}$, then

$$|F(x) - \pi(x)| \leq \pi(x) + F(x) \leq \lambda_s(A^c) + \mu(A^c).$$

Otherwise, put $y = (x_1 \wedge (1-a), \ldots, x_s \wedge (1-a))$. Then $y \in A$ and

$$\begin{aligned} |F(x) - \pi(x)| &\leq |F(y) - \pi(y)| + |\pi(x) - \pi(y)| + |F(x) - F(y)| \\ &\leq \epsilon + \lambda_s(A^c) + \mu(A^c). \end{aligned}$$

An easy computation gives

$$\begin{aligned} \mu(A) &= \mu([0, 1-a]^s) - \mu([0, 1-a]^s \backslash A) \\ &\leq (1-a)^s - \epsilon - s[\epsilon + a(1-a)^{s-1}], \end{aligned}$$

and we conclude that

$$\|F - \pi\|_\infty \leq (2+s)\epsilon + (1 - (1-2a)^s) + (1 - (1-a)^s) + sa(1-a)^{s-1}.$$

(c) The assertion of the lemma now follows readily from (a) and (b). □

A.5 Uniformly Distributed Sequences and Discrepancy

Definition A.5.1 A sequence $\xi = (\xi_n)_{n\geq 1}$ in $[0,1]^s$ is called *uniformly distributed* (u.d.) in $[0,1]^s$ if for every $x = (x_1, \ldots, x_s)$ in $[0,1]^s$,

$$\lim_{N\uparrow\infty} \frac{1}{N} \sum_{n=1}^{N} \prod_{i=1}^{s} 1_{\xi_n^i \leq x_i} = \pi(x).$$

The rate of convergence can be measured by several *discrepancies*:

$$D_\infty^*(\xi, N) = \sup_{x\in[0,1]^s} \left| \pi(x) - \frac{1}{N} \sum_{n=1}^{N} 1_{[[0,x]]}(\xi_n) \right|$$

$$D_\infty(\xi, N) = \sup_{x,y\in[0,1]^s x\leq y} \left| \prod_{i=1}^{s} (y_i - x_i) - \frac{1}{N} \sum_{n=1}^{N} 1_{[[x,y]]}(\xi_n) \right|$$

$$D_p^*(\xi, N) = \left\| \pi(x) - \frac{1}{N} \sum_{n=1}^{N} 1_{[[0,x]]}(\xi_n) \right\|_{L^p(dx)} \qquad 1 \leq p < \infty$$

$$D_p(\xi, N) = \left\| 1_{\{x\leq y\}} \left(\prod_{i=1}^{s} (y_i - x_i) - \frac{1}{N} \sum_{n=1}^{N} 1_{[[x,y]]}(\xi_n) \right) \right\|_{L^p(dx\,dy)} \qquad 1 \leq p < \infty.$$

It is easily checked that, for every ξ,

$$\frac{1}{2N} \leq D_\infty^*(\xi, N) \leq 1$$

and

$$D_\infty^*(\xi, N) \leq D_\infty(\xi, N) \leq 2^s D_\infty^*(\xi, N).$$

Theorem A.5.2 The following statements are equivalent:

(i) The sequence ξ is u.d. in $[0,1]^s$.

(ii) For every p in $[1,\infty]$, $D_p^*(\xi, N) \to 0$ as $N \to \infty$.

(iii) There exists p in $[1,\infty]$ such that $D_p^*(\xi, N) \to 0$ as $N \to \infty$.

(iv) The probability measure $\mu_N = (1/N) \sum_{n=1}^{N} \delta_{\xi_n}$ converges narrowly to the Lebesgue measure λ_s as $N \to \infty$.

Proof (i) $\Rightarrow$ (ii). If ξ is u.d., the distribution function F_N of μ_N converges pointwise to π. By Lebesgue's theorem on dominated convergence, F_N converges in $L^1(dx)$, and also uniformly by Lemma A.4.3.

(iii) $\Rightarrow$ (iv). It also follows from Lemma A.4.3 that if $D_p^*(\xi, N) \to 0$ then $F_N \to \pi$ uniformly. Therefore $\int g d\mu_N \to \int g d\lambda_s$ for every finite sum g of indicator functions of s-dimensional intervals. Since every continuous function can be uniformly approximated by such functions, we get (iv).

(iv) $\Rightarrow$ (i). The boundaries of s-dimensional intervals have Lebesgue measure zero, so their indicator functions are Riemann integrable. □

Proposition A.5.3 (Hermann Weyl 1926) The following statement is also equivalent to conditions (i) to (iv) above:

(v) For every $m \neq (0, \ldots, 0)$ in $\mathbf{Z}^s$,

$$\frac{1}{N} \sum_{n=1}^{N} \exp(2i\pi\langle m, \xi_n \rangle) \to 0 \quad \text{as} \quad N \to \infty.$$

Proof (iv) $\Rightarrow$ (v). The functions $x \rightsquigarrow \exp(2i\pi\langle m, x\rangle)$ are bounded and continuous on $[0, 1]^s$.

(v) $\Rightarrow$ (iv). We have $\mu_N(f) \to \lambda_s(f)$ for every linear combination f of exponentials. These functions generate an algebra, which is closed under complex conjugation, contains the constant functions, and separates points in the s-dimensional torus T_s. By the Stone–Weierstrass theorem, the assertion is true for every f in $\mathscr{C}_b(T_s)$, for every Riemann integrable function on T_s, and therefore for every bounded continuous function on $[0, 1]^s$. □

A.5.4 Examples

(1) Let $\alpha_1, \ldots, \alpha_s$ be irrational numbers such that $1, \alpha_1, \ldots, \alpha_n$ are linearly independent over the rationals. If we denote by $\{x\}$ the fractional part of the real number x, then the sequence $(\{n\alpha_1\}, \ldots, \{n\alpha_s\})$ is u.d. in $[0, 1]^s$.

(2) Let $(p_n)_{n \in \mathbf{N}}$ be the increasing sequence of prime numbers. For every irrational α, the sequence $(\{\alpha p_n\})$ is u.d. in $[0, 1]$.

(3) The sequence $(\{\sqrt{n}\})$ is u.d. in $[0, 1]$.

(4) If α is irrational, the sequence $(\{\alpha n^s\}, \{\alpha(n+1)^s\}, \ldots, \{\alpha(n+s-1)^s\})$ is u.d. in $[0, 1]^s$.

(5) For λ_1-a.e. $\theta > 1$, the sequence $(\{\theta^n\})$ is u.d. in $[0, 1]$ (Koksma).

(6) For λ_1-a.e. $k > 0$, the sequence $(\{k(\frac{3}{2})^n\})$ is u.d. in $[0, 1]$.

A.5.5 Extension to μ-Distribution

The notion of uniform distribution can be extended to a more general setting. Let μ be a probability measure on a Polish space $(E, \mathscr{E})$. A sequence $\xi = (\xi_n)$ in E is called μ-*distributed* if the sequence of measures $(1/N) \sum_{n=1}^{N} \delta_{\xi_n}$ converges narrowly to μ.

For instance, inside every nonempty interval of $\mathbb{R}_+$, there exist $2^{\aleph_0}$ numbers x such that $(\{x(\frac{3}{2})^n\})$ is μ-distributed, with $\mu = [k/(k+1)]\lambda_1 + [1/(k+1)]\delta_0$ and $k = \log 2/\log \frac{3}{2}$ [G. Choquet (1)].

B THE MONTE CARLO METHOD

The basic idea of the Monte Carlo method is to use pseudo-random sequences to simulate random behavior and then apply the strong law of large numbers in an appropriate fashion. Quasi-Monte Carlo methods, which we will study in Section C, use sequences that satisfy only the weaker property of equidistribution. At the end of Section C, and later in Chapters 3 and 5, we will compare the different areas of application, rates of convergence, and stopping criteria for the different methods.

Obtaining sequences of real numbers that imitate perfectly the behavior of random sequences is not only difficult, as can be seen by the history of the successive attempts [Dellacherie, 1978], but logically impossible [Martin-Löf, 1966]. Nevertheless, the art of constructing pseudo-random sequences is at now very advanced (see the references in Section C.5.3) and yields excellent generators which are generally easy to implement.

> Ich bin Zarathustra, der Gottlose: ich Koche mir noch jeden Zufall in *meinem* Topfe. Und erst, wenn er da gar gekocht ist, heisze ich ihn willkommen, als *meine* Speise.[1] [Friedrich Nietzsche, 1883]

B.1 Algorithms for Simulation

Before we can use the Monte Carlo method to compute the expectation of a random variable F, we must first define a random variable with the same law as F on the probability space $([0,1]^m, \mathscr{B}([0,1]^m), dx^m)$ for some m in $\mathbf{N} \cup \{+\infty\}$. Simulation algorithms are techniques that yield a random variable with a given law on $([0,1]^m, \mathscr{B}([0,1]^m), dx^m), m \in \mathbf{N} \cup \{+\infty\}$. We can rephrase this in the following language: A *Borel sequence* is a sequence $(U_n)_{n\geq 1}$ of independent random variables, where each U_n takes

[1] I am Zarathustra the godless: I cook every hazard in *my* pot, and when it is well done, I welcome it as *my* dish.

values in $[0, 1]^s$ with law dx^s. How can we obtain from $(U_n)_{n\geq 1}$ a sequence of independent variables each with the same law as F?

There are general procedures to solve this problem, such as the inversion, composition, and rejection methods, as well as more specialized algorithms for specific cases. It is beyond the scope of this book to give even an account of these techniques for which we refer to Rubinstein [1981], Morgan [1984], Ripley [1987], and especially to the reference book of L. Devroye [1986].

Let us emphasize the fact that these techniques, as we shall see below, can also be used with quasi-Monte Carlo methods instead of the Monte Carlo method (as long as the copy of F is regular enough to be Riemann-integrable, see the discussion in Section C.5.3 and Chapter 3, Section A.2).

We restrict our attention to the *rejection method* [Von Neumann, 1951], which is sometimes called the *acceptance method*. Let X be a random variable with values in $\mathbb{R}^d$ and density function f. Assume that we know another distribution function g, called the *envelope*, from which it is easy to sample, and that there exists $k > 1$ such that $f \leq kg$. Then the random variable X can be simulated in the following way.

(i) We simulate a random variable H with density g.

(ii) Independently, we generate a variable U with uniform distribution on [0,1]

- if $kg(H)U \leq f(H)$ set $Y = H$
- if $kg(H)U > f(H)$ return to **(i)** and repeat, independently of what happened before.

It can now be proved that f is a density function for Y. Indeed, the point P with coordinates $P_1 = H, P_2 = kg(H)U$ is uniformly distributed under the graph of kg. By the following lemma, if we restrict P to the cases when it lies below the graph of f, its distribution becomes uniform below this graph and therefore the distribution of P_1 has density function f.

Lemma B.1.1 Let $(P_n)_{n\geq 1}$ be a sequence of independent random variables with the same distribution μ on $\mathbb{R}^k$. Let A be a Borel set in $\mathbb{R}^k$ such that $\mu(A) > 0$, and let $T = \inf\{n \geq 1 : P_n \in A\}$. Then P_T has distribution $(1/\mu(A))\mu|_A$ and $\mathbb{E}(T) = 1/\mu(A)$.

Proof Exercise.

B.1.2 Exercise

By slightly modifying the rejection method, show that it is possible to simulate a distribution with density function f even if f is known only by approximations $\underline{f_n}$ and $\overline{f_n}$ such that for any n, $\underline{f_n} \leq f \leq \overline{f_n}$, and the sequence $\underline{f_n} - \overline{f_n}$ tends pointwise to zero. This property allows a wide field of application for the rejection method.

B.1.3 Exercise

A generator of pseudo-random numbers is a device that provides sequences of numbers imitating Borel sequences. Most generators are built on congruence methods. Let us assume that a sequence of such generators is given, and that, for every k, the kth generator, when used N_k times, supplies the numbers $\{n/N_k; n = 0, \ldots, N_k - 1\}$ in an arbitrary order. As soon as the kth generator has been used N_k times, we move to the next generator, and we record all the results one after the other to create a sequence. Show that if the sequence $(N_k)_{k\geq 1}$ is increasing and satisfies

$$\lim_{k\uparrow\infty} \frac{N_{k+1}}{N_1 + \cdots + N_k} = 0$$

then the resulting sequence is uniformly distributed in [0,1].

B.1.4 Exercise: Wiener–Siegel Polychotomy Algorithm

In their interpretation of quantum mechanics, N. Wiener and A. Siegel [1953, 1955] consider the reduction process of the wave packet to be a random sampling, that is, a simulation. Let (Ψ_i) be the sequence of eigenfunctions of the observables and, for $\sum_i |a_i|^2 = 1$, let $\Psi = \sum_i a_i \Psi_i$ be the unit vector that represents the wave function. Show that we can simulate a random variable N with distribution $\sum_{n\in\mathbf{N}} |a_n|^2 \delta_n$ as follows: first draw a sequence of bivariate normalized Gaussian variables (g_n) and then select the index n that achieves the maximum of $|a_n|/|g_n|$.

B.2 Glivenko–Cantelli Theorem

Due to its wide usefulness, the following theorem is sometimes called the *fundamental theorem of statistics*.

Theorem B.2.1 Let $(U_n)_{n\geq 1}$ be a Borel sequence in $[0, 1]^s$. We define the empirical distribution function of (U_n) by

$$\pi_N(x) = \frac{1}{N}\sum_{n=1}^{N} \mathbf{1}_{[[0,x]]}(U_n), \qquad N \geq 1.$$

(Recall the notation $\pi(x) = \lambda_s([[0,x]])$.) Then,

(i) a.s., the empirical distribution

$$\mu_N = \frac{1}{N}\sum_{n=1}^{N} \delta_{U_n}$$

converges narrowly to λ_s;

(ii)

$$\sup_{x\in[0,1]^s} |\pi_N(x) - \pi(x)| \to 0 \quad \text{a.s.} \quad \text{as} \qquad N \to \infty.$$

Proof (i) For any countable dense set D in $\mathscr{C}([0,1]^s, \mathbb{R})$,

$$\mu_N(g) \to \lambda_s(g) \qquad \text{a.s.} \quad \text{for any} \quad g \in D,$$

by the strong law of large numbers and this is enough to ensure narrow convergence.

(ii) Immediate from (i) by using Theorem A.5.2. □

Theorem B.2.2 Let $(U_n)_{n\geq 1}$ be a Borel sequence in [0,1], let F be a distribution function on $\mathbb{R}$, and let G be the inverse function of F, that is,

$$G(y) = \inf\{x: F(x) \geq y\} \qquad 0 \leq y \leq 1.$$

Set

$$F_N(x) = \frac{1}{N}\sum_{n=1}^{N} 1_{\{G(U_n)\leq x\}}$$

Then

(i) a.s.

$$\lim_{N\uparrow\infty} \sup_F \sup_x |F_N(x) - F(x)| = 0$$

(ii) if F is a continuous function, the distribution of

$$\sup_x |F_N(x) - F(x)|$$

does not depend on F.

Proof (i) By using B.2.1., for almost every $\omega \in \Omega$,

$$\frac{1}{N}\sum_{n=1}^{N} 1_{\{U_n(\omega)\le F(x)\}} \to F(x) \qquad \text{as} \quad N \uparrow \infty$$

uniformly with respect to F and x. Since $G(b) \le x$ iff $b \le F(x)$. we have

$$\frac{1}{N}\sum_{n=1}^{N} 1_{\{G(U_n(\omega))\le x\}} \to {}_{N\uparrow\infty} F(x).$$

(ii) If F is continuous it takes all the values in the interval $]0, 1[$, and thus

$$\sup_{x\in\mathbb{R}} \left| \frac{1}{N}\sum_{n=1}^{N} 1_{\{U_n(\omega)\le F(x)\}} - F(x) \right|$$
$$= \sup_{y\in[0,1]} \left| \frac{1}{N}\sum_{n=1}^{N} 1_{\{U_n(\omega)\le y\}} - y \right| \quad \text{a.s.} \qquad \square$$

B.2.3 Exercise

Let μ be a probability measure on $\mathbb{R}$ with continuous distribution function F, and let F_N be as defined above. Show that the distribution of $\|F_N - F\|_{L^p(\mu)}$, for $1 \le p < \infty$, is independent of μ.

B.3 Central Limit Theorem for Empirical Distributions

Let $(U_n)_{n\ge1}$ be a Borel sequence in $[0, 1]^s$, and consider the random variables

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{[[0,x]]}(U_i).$$

(a) We know from the central limit theorem that for any fixed x, $\sqrt{n}(F_n(x) - \pi(x))$ converges in law to $N(0, \pi(x) - (\pi(x))^2)$.

(b) Consider now the stochastic process

$$\sqrt{n}(F_n(x) - \pi(x)) \qquad x \in [0,1]^s.$$

By the k-dimensional central limit theorem, the random vector $\{\sqrt{n} \cdot (F_n(x_1) - \pi(x_1)), \ldots, \sqrt{n}(F_n(x_k) - \pi(x_k))\}$ converges in law to $N_k(0, \Sigma)$ for any $x_1, \ldots, x_k$ in $[0,1]^s$, where

$$\Sigma = \mathbb{E}(HH^t) - \mathbb{E}H.\mathbb{E}H^t$$

$$H = \begin{pmatrix} 1_{[[0,x_1]]}(U_1) \\ \vdots \\ 1_{[[0,x_k]]}(U_1) \end{pmatrix}$$

(c) If $s = 1$, these limit distributions are the finite-dimensional distributions of a real centered Gaussian process Z_t with covariance

$$\mathbb{E}Z_t Z_s = s \wedge t - st \qquad s, t \text{ in } [0,1].$$

The Brownian bridge $B_t - tB_1$ has exactly this covariance function. Since the distribution of a centered Gaussian process is entirely determined by its covariance function, we can recognize Z_t as a Brownian bridge. This makes it easier to understand the following Doob–Donsker form of the Kolmogorov–Smirnov theorem.

Theorem B.3.1 Let $(U_n)_{n\geq 1}$ be a Borel sequence in $[0,1]^s$ and let (F_n) be the sequence of associated empirical distribution functions. We set

$$D_\infty^*(n) = \sup_{x\in[0,1]^s} |F_n(x) - \pi(x)|$$

$$D_\infty^+(n) = \sup_{x\in[0,1]^s} (F_n(x) - \pi(x))$$

$$D_\infty^-(n) = \sup_{x\in[0,1]^s} (\pi(x) - F_n(x)).$$

(i) Case $s = 1$. Let $(B_x)_{x\in[0,1]}$ be a standard one-dimensional Brownian motion. We have the following convergences in law

$$\sqrt{n}D_\infty^*(n) \xrightarrow{\mathcal{L}} \sup_{x\in[0,1]} |B_x - xB_1|$$

$$\sqrt{n}D_\infty^+(n) \xrightarrow{\mathcal{L}} \sup_{x\in[0,1]} (B_x - xB_1)$$

$$\sqrt{n}D_\infty^-(n) \xrightarrow{\mathcal{L}} \sup_{x\in[0,1]} (xB_1 - B_x).$$

The limit distribution functions are given for $\lambda > 0$ by

$$\mathbb{P}\left(\sup_{x\in[0,1]} |B_x - xB_1| \le \lambda\right) = \sum_{j=-\infty}^{+\infty} (-1)^j e^{-2j^2\lambda^2}$$

$$\mathbb{P}\left(\sup_{x\in[0,1]} (B_x - xB_1) \le \lambda\right) = \mathbb{P}\left(\sup_{x\in[0,1]} (xB_1 - B_x) \le \lambda\right)$$
$$= 1 - e^{-2\lambda^2}$$

(ii) Case $s > 1$. There exists a Gaussian centered real process $(Z_x)_{x\in[0,1]^s}$ with continuous paths and covariance

$$\mathbb{E}[Z_x Z_y] = \pi(x \wedge y) - \pi(x)\pi(y) \qquad x, y \in [0,1]^s$$

such that the random variables $\sqrt{n}D_\infty^*(n)$, $\sqrt{n}D_\infty^+(n)$, and $\sqrt{n}D_\infty^-(n)$ converge in law to

$$\sup_{x\in[0,1]^s} |Z_x|, \quad \sup_{x\in[0,1]^s} Z_x, \quad \text{and} \quad \sup_{x\in[0,1]^s} (-Z_x),$$

respectively.

B.3.2 Comments

Let $s = 1$, and consider a random sample from a continuous distribution function F, with empirical distribution function F_N. From part (ii) of Theorem B.2.2, we may infer that the conclusions of part (i) of Theorem B.3.1 are still valid with the same limit distributions.

We give a rough sketch of the proof of Theorem B.3.1 in the case $s = 1$. We have already noted that finite-dimensional distributions of the process

$\{\alpha_n(x) = \sqrt{n}(F_n(x) - \pi(x)), x \in [0,1]\}$ converge to those of a Brownian bridge.

Consider the compact metrizable space $\Omega: [\overline{\mathbb{R}}]^{\mathbb{Q}\cap[0,1]}$ (where $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$), equipped with its Borel σ-field. It follows from the Stone–Weierstrass theorem that the set of continuous functions on Ω, which depend only on a finite number of coordinates, is a dense subset of $\mathscr{C}(\Omega, \mathbb{R})$. The probability distribution p_n of $\{\alpha_n(x), x \in [0,1]\}$ converges on these functions to the probability distribution p of $\{B_x - xB_1, x \in [0,1]\}$, and this proves narrow convergence of p_n to p. But the functional on Ω

$$\phi(\omega) = \sup_{x \in \mathbb{Q}\cap[0,1]} |\omega(x)|$$

is not p-Riemann integrable: it is continuous in the topology of uniform convergence, but not in that of pointwise convergence. So we have to change topology on Ω to make ϕ continuous. In fact, there are now far more continuous bounded functions on Ω, and the convergence of finite-dimensional distributions is no longer sufficient to ensure that p_n converges narrowly to p. We need the sequence (p_n) to have the property of relative compactness, which is involved by Prokhorov's lemma. The uniqueness of its limit point follows from the convergence of finite-dimensional distributions, and so p_n converges to p.

B.3.3 To Go Further

More precise results than those of Theorem B.3.1 can be stated. For $s = 1$, there is actually a strong invariance principle for empirical distribution functions that allows strong approximation of the process $\{\sqrt{n}(F_n(x) - F(x)), x \in [0,1]\}$ by a Gaussian process $\{Z_x, x \in [0,1]\}$, provided F is continuous. In the particular case $F(x) = x$, Komlos, Major, and Tusnady [1975] have exhibited a sequence of independent Brownian motions $(B^n)_{n\geq 1}$ such that

$$\sup_{x\in[0,1]} \left| \sqrt{n}(F_n(x) - x) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (B_x^n - xB_1^n) \right| = O\left(\frac{(\log n)^2}{\sqrt{n}} \right) \quad \text{a.s.}$$

Recently, F. Bonvalot [1991] and N. Castelle [1991] have completed their demonstration by computing three constants C, K, and λ such that for every $n \geq 3$ and $y \geq 0$,

$$\mathbb{P}\left(\sup_{1\le m\le n} \sup_{0\le x\le 1} \left| m(F_m(x) - x) - \sum_{i=1}^{m} (B_x^i - xB_1^i) \right| > (y + C\log n)\log n \right)$$
$$\le K\exp(-\lambda y).$$

For $s = 2$, Tusnady [1977] has obtained a strong invariance principle with the same estimate $O((\log n)^2/n)$; higher dimensions remain as yet unexplored.

B.4 Stopping Criteria and Asymptotic Rate

The stopping criteria in the classical Monte Carlo method have to be probabilistic. Typically, if n is sufficiently large, there is only a small probability that the n-step computed result is far away from the requested value.

> A mesure que l'événement observé se répète, cette vraisemblance augmente et finirait par se confondre avec la certitude, si le nombre des répétitions devenait infini. Il y a ici deux sortes d'approximations: l'une d'elles est relative aux limites prises, de part et d'autre, des possibilités qui donnent au passé plus de vraisemblance: l'autre approximation se rapporte à la probabilité que ces possibilités tombent dans ces limites. La répétition de l'événement composé accroît de plus en plus cette probabilité, les limites restant les mêmes: elle resserre de plus en plus l'intervalle de ces limites, la probabilité restant la même: dans l'infini, cet intervalle devient nul, et la probabilité se change en certitude.[2] [Pierre Simon Laplace, 1821]

Suppose that $(X_n)_{n\ge 1}$ is a sequence of i.i.d. random variables, and $S_n = X_1 + \cdots + X_n$. If $E(X_1^2) < \infty$, for all $\lambda > 0$,

$$\mathbb{P}\left(\left| \frac{S_n}{n} - \mathbb{E}X_1 \right| \ge \lambda \right) \le \frac{\mathrm{Var}\,(X_1)}{n\lambda^2}.$$

Let

[2]As the observed event is repeated, this likelihood correspondingly increases, and would become certainty if the number of repetitions became infinite. There are two sorts of approximation here: one, with respect to given limits, above and below, is of the possibilities that give more likelihood to past events; the other approximation concerns the probability that these possibilities fall within these limits. The repetition of the event steadily increases this probability, for fixed limits: it steadily narrows the interval of these limits, for fixed probability: in the limit, this interval has zero length, and probability becomes certainty.

$$\phi(t) = \mathbb{E}(e^{tX_1}), \qquad t \in \mathbb{R}.$$

The set $\{t: \phi(t) < \infty\}$ is an interval I containing 0, and the Cramer transform of the distribution of X_1 is defined by

$$h(b) = \sup_{t \in I} \{bt - \log \phi(t)\}.$$

It is easy to see that $\psi(t) = \log \phi(t)$ is a convex function on I and if the interior $\overset{o}{I}$ of I is nonempty, then ψ is differentiable on $\overset{o}{I}$. Moreover, if $0 \in \overset{o}{I}$, then $E|X_1| < \infty$; setting $m = E(X_1)$, we have $\psi'(0) = m$ (see Dacunha-Castelle and Duflo [1983] for details). We can now state Chernov's theorem, which is the starting point of the modern theory of large deviations.

Theorem B.4.1 Let us assume there exists $t_0 > 0$ such that $E(e^{t_0|X_1|}) < \infty$. Then for $b > E(X_1)$,

$$\mathbb{P}(S_n \geq nb) \leq e^{-nh(b)},$$

and for $b < E(X_1)$,

$$\mathbb{P}(S_n \leq nb) \leq e^{-nh(b)}.$$

In both cases, $h(b) > 0$.

Proof Suppose $b > E(X_1)$; we handle the other case by replacing (X_n) with $(-X_n)$. We have

$$\begin{aligned}\mathbb{P}(S_n \geq nb) &\leq \inf\{\mathbb{E}(e^{-unb+uS_n}), u \in I \cap]0, \infty[\} \\ &= \inf\{e^{-n(bu-\psi(u))}, u \in I \cap]0, \infty[\}\} \\ &= \exp(-n \sup\{bu - \psi(u), u \in I \cap]0, \infty[\}).\end{aligned}$$

Set $g(u) = bu - \psi(u)$ for $u \in I$. Then g is concave, $g(0) = 0$, and $g'(0) = b - E(X_1) > 0$. It follows that

$$\sup\{g(u), u \in I \cap]0, \infty[\} = h(b) > 0. \qquad \square$$

The asymptotic speed is ruled by the Hartman–Wintner law of the iterated logarithm: If the random variables (X_n) are i.i.d., centered, with variance $\sigma^2 < \infty$, then

$$\overline{\lim} \frac{S_n}{\sigma\sqrt{2n \log \log n}} = -\underline{\lim} \frac{S_n}{\sigma\sqrt{2n \log \log n}} = 1 \qquad \text{a.s.}$$

Let F_n be the empirical distribution function of a Borel sequence in $[0, 1]^s$. For all x in $[0, 1]^s$ we have

$$\overline{\lim} \sqrt{\frac{n}{2 \log \log n}} \, (F_n(x) - \pi(x)) = \sqrt{\pi(x)(1 - \pi(x))} \qquad \text{a.s.}$$

and hence

$$\overline{\lim} \sqrt{\frac{n}{2 \log \log n}} \, D_\infty^*(n) \geq \sup_x \sqrt{\pi(x)(1 - \pi(x))} = \frac{1}{2} \qquad \text{a.s.}$$

The last inequality is actually an equality [Chung, 1949; Kiefer, 1961] as we state in the following theorem.

Theorem B.4.2 Let (U_n) be a Borel sequence in $[0, 1]^s$. Then

$$\overline{\lim} \sqrt{\frac{n}{2 \log \log n}} \, D_\infty^*(n) = \frac{1}{2},$$

and the same equalities hold for $D_\infty^+(n)$ and $D_\infty^-(n)$ in place of $D_\infty^*(n)$.

B.4.3 Exercise

Let F_n be the empirical distribution functions of a Borel sequence in $[0, 1]^s$. For $1 \leq p < \infty$, set

$$D_p^*(n) = \|F_n - \pi\|_{L^p([0,1]^s, \lambda_s)}.$$

(1) Deduce from Theorem B.4.2 that

$$\overline{\lim}\ \sqrt{\frac{n}{2\log\log n}}\ D_p^*(n) \le \|\sqrt{\pi(x)(1-\pi(x))}\|_p \qquad \text{a.s.}$$

(2) Show that

$$\mathbb{E}[(D_2^*(n))^2] = \frac{1}{n}\left(\frac{1}{2^s} - \frac{1}{3^s}\right).$$

B.4.4 Remarks

Let (X_n) be real valued i.i.d. random variables with distribution μ and continuous distribution function F and let (F_n) be the associated empirical distribution functions. The following are true (cf. Exercise B.2.3).

(1) [Finkelstein 1971]

$$\overline{\lim}\ \sqrt{\frac{n}{2\log\log n}}\ \|F_n - F\|_{L^2(\mu)} = \frac{1}{\pi} \qquad \text{a.s.}$$

(2) [Mogul'skii 1977]

$$\underline{\lim}\ \sqrt{2n\log\log n}\|F_n - F\|_{L^2(\mu)} = \frac{1}{2} \qquad \text{a.s.}$$

$$\underline{\lim}\ \sqrt{2n\log\log n}\|F_n - F\|_\infty = \frac{\pi}{2} \qquad \text{a.s.}$$

B.4.5 Exercise

Suppose that $D_\infty^*(n)$ and $D_2^*(n)$ are calculated relative to a sampling from the uniform distribution on $[0, 1]$, and that $\tilde{D}_\infty^*(n)$ and $\tilde{D}_2^*(n)$ are calculated relative to a second (independent) sampling from the same distribution. Show that, for every $\epsilon > 0$, a.s. there exists n_0 such that, for $n \ge n_0$,

$$\left|\log\frac{D_\infty^*(n)}{\tilde{D}_\infty^*(n)}\right| \le \log\log\log n - \log\frac{\pi}{2} + \epsilon$$

and

$$\left|\log\frac{D_2^*(n)}{\tilde{D}_2^*(n)}\right| \le \log\log\log n + \log\frac{4}{\pi} + \epsilon.$$

B.5 Rate of Convergence of Stochastic Approximation Algorithms

Suppose that we are given a random variable $f(x)$ that depends on a parameter x, and we want to approximate the point x^* where the average $F(x) = \mathbb{E}(f(x))$ assumes a given value, or where the maximum of F occurs. Stochastic approximation procedures are especially useful in this situation: we use the Robbins–Monro procedure [1951] for the first problem and the Kiefer–Wolfowitz procedure [1952] for the second.

In the Robbins–Monro scheme for finding some x^* that meets $F(x) = \beta$, we start with an initial guess X_1. We make successive observations (Y_{n+1}), and (X_n) is chosen according to the formula

$$X_{n+1} = X_n + \gamma_n(\beta - Y_{n+1}).$$

Under appropriate hypotheses, X_n converges to x^*. (See Figure 2.1.)

Theorem B.5.1 Consider the recursive equations

$$X_{n+1} = X_n + \gamma_n(\beta - Y_{n+1}) \qquad n \geq 1,$$

where X_n and Y_n are $\mathbb{R}^d$-valued random variables, $\beta \in \mathbb{R}^d$, and X_1 belongs to any $L^p(\Omega, \mathbb{R}^d)$, $1 \leq p < \infty$. Assume that the following three conditions hold:

(H_1) (γ_n) is a decreasing sequence of positive numbers chosen so that $\sum \gamma_n = \infty$ and $\sum \gamma_n^2 < \infty$.

(H_2) There exists $b > 0$ such that $|Y_n| \leq b$ for all $n \geq 2$.

(H_3) Let $\mathcal{F}_1 = \sigma(X_1)$ and $\mathcal{F}_n = \sigma(X_1, \ldots, X_n, Y_2, \ldots, Y_n)$, for $n \geq 2$. There exists a Borel function $F : \mathbb{R}^d \to \mathbb{R}^d$ satisfying

$$\mathbb{E}[Y_{n+1} | \mathcal{F}_n] = F(X_n)$$

and a point x^* in $\mathbb{R}^d$ such that, for all δ in $]0, 1[$,

$$\inf_{\delta \leq |x - x^*| \leq 1/\delta} < x - x^*, F(x) - \beta > > 0.$$

Then X_n converge to x^* a.s. and in $L^p(\Omega; \mathbb{R}^d)$, for all p in $[1, \infty[$.

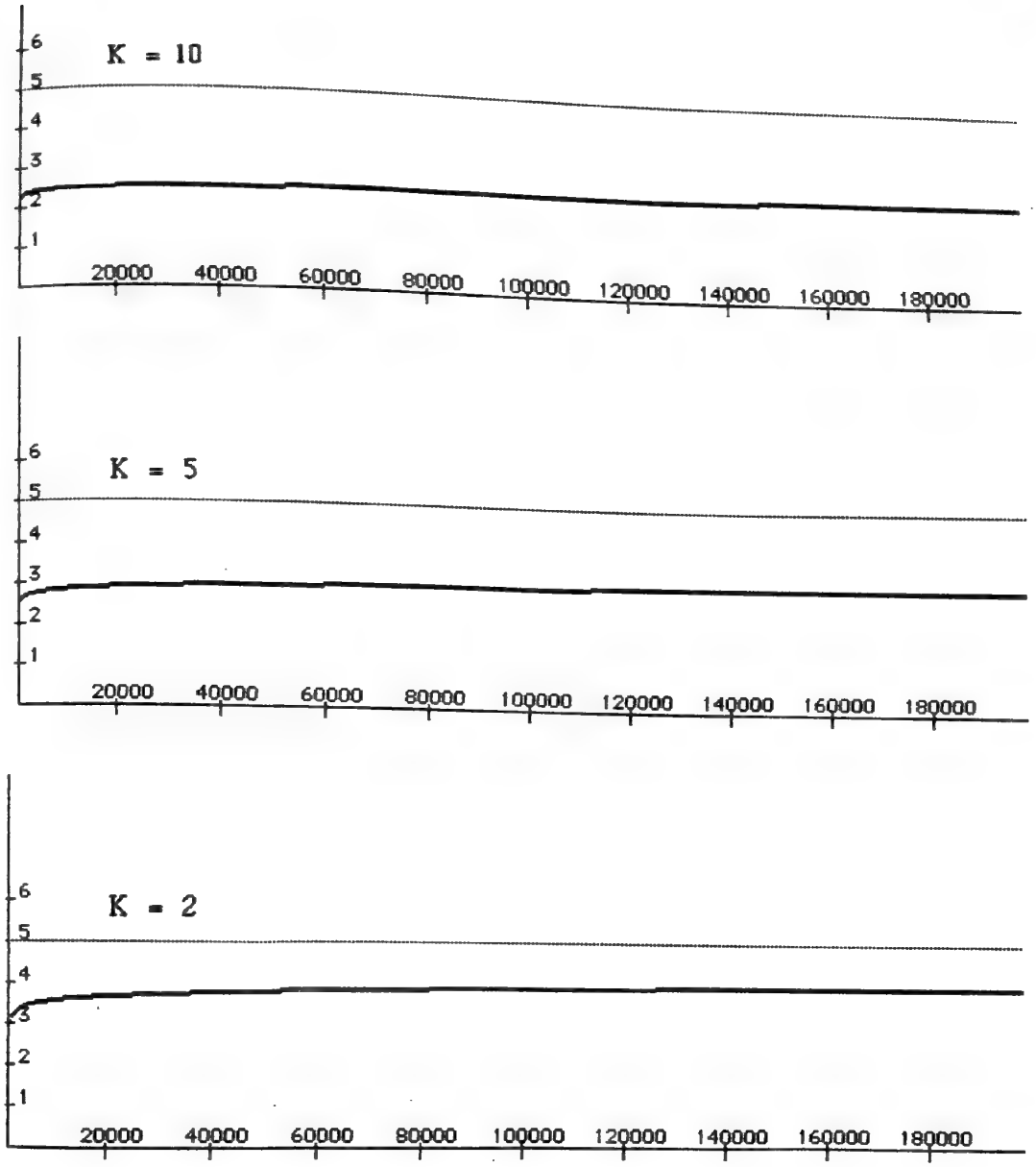


Figure 2.1 Simulation of the Robbins–Monro algorithm

$$X_{n+1} = X_n + \frac{2}{5+Kn}\left(\tfrac{1}{2} - 1_{\{X_n > 10.U_{n+1}\}}\right)$$

Starting point $X_1 = 2$, from $n = 1$ to 200,000, and for $K = 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01, 0.005$, and 0.001.

Proof Let

$$s = \sum_{n=1}^{\infty} \gamma_n^2 \qquad \text{and} \qquad T_n = X_n - x^*, \qquad \text{for} \quad n \geq 1.$$

(1) *Almost sure convergence.* We have

$$\text{(B.5.2)} \quad \mathbb{E}[|T_{n+1}|^2 | \mathcal{F}_n] = |T_n|^2 + 2\gamma_n \langle \beta - F(X_n), T_n \rangle + \gamma_n^2 \mathbb{E}[|\beta - Y_{n+1}|^2 | \mathcal{F}_n]$$

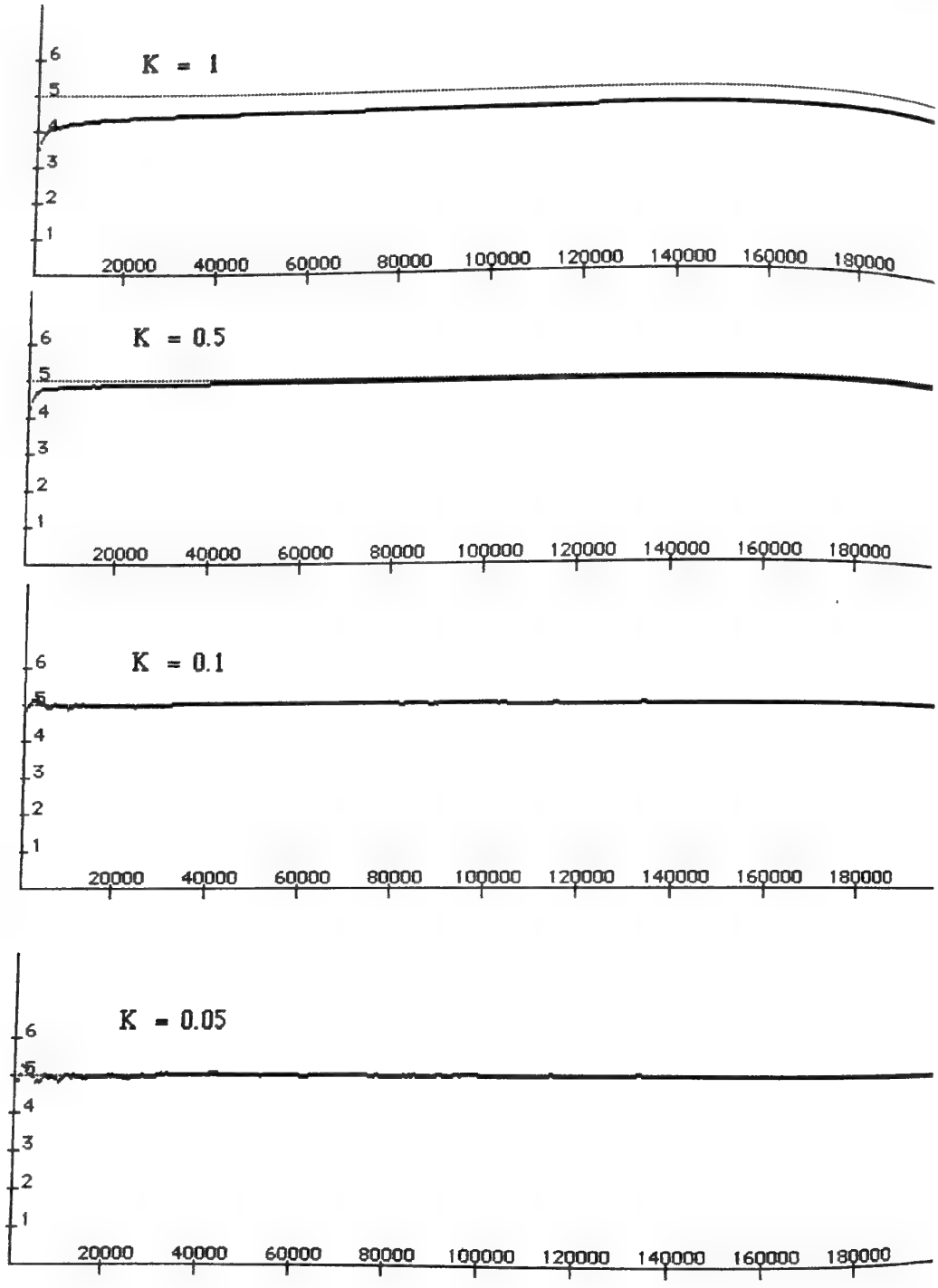


Figure 2.1 (*Continued*)

It follows from (H_1) and (H_2) that

$$\sum_{k=1}^{n} \gamma_k^2 \mathbb{E}[|Y_{k+1} - \beta|^2 | \mathcal{F}_k] \leq s(|\beta| + b)^2.$$

For $n \geq 1$, let

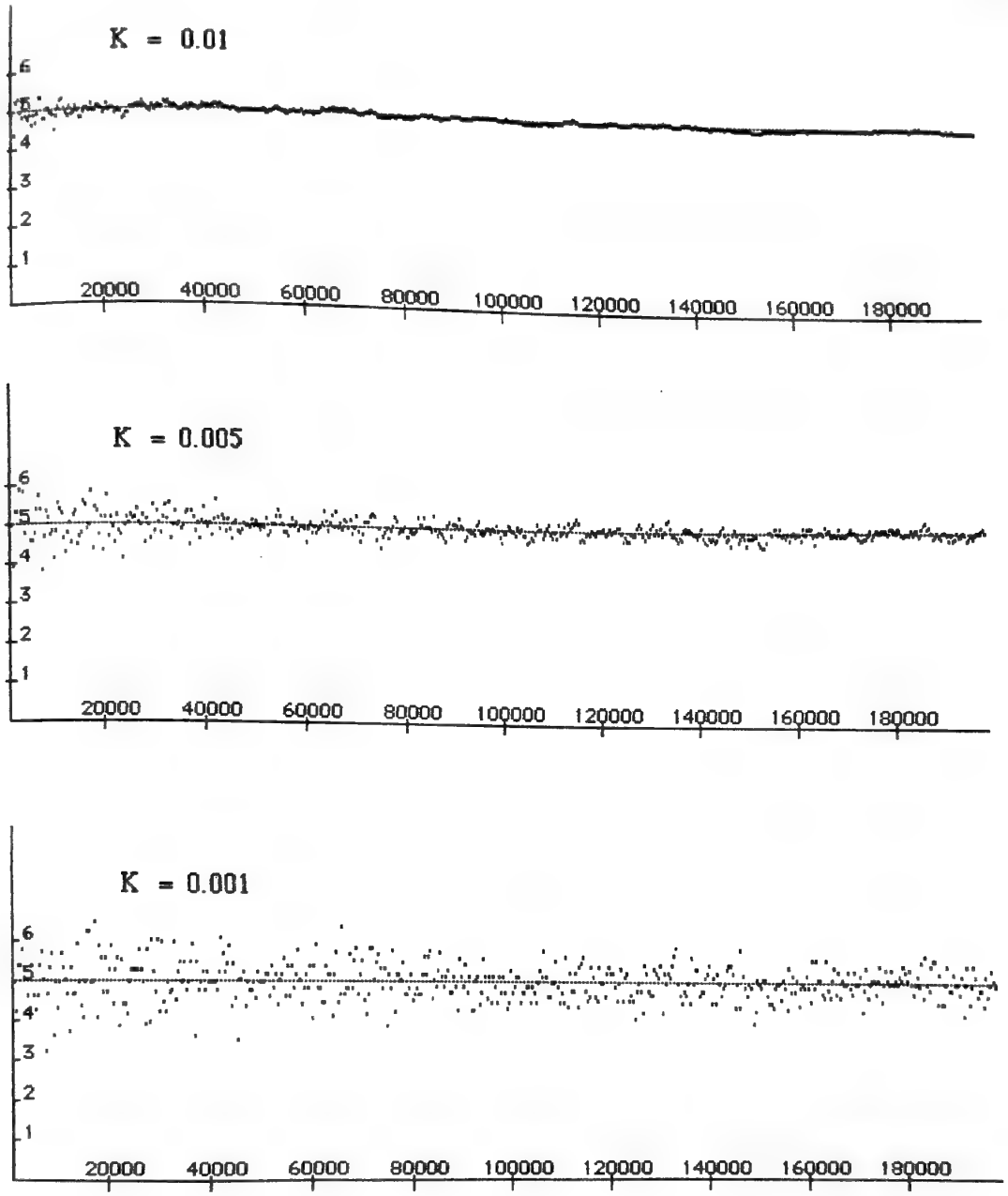


Figure 2.1 (*Continued*)

$$Z_n = |T_n|^2 + s(|\beta| + b)^2 - \sum_{k=1}^{n-1} \gamma_k^2 \mathbb{E}[|Y_{k+1} - \beta|^2 | \mathcal{F}_k].$$

Then

$$\mathbb{E}[Z_{n+1} | \mathcal{F}_n] = Z_n + 2\gamma_n \langle \beta - F(X_n), T_n \rangle,$$

and by (H_3) it follows that (Z_n) is a nonnegative supermartingale. Thus (Z_n) converges a.s. as n tends to infinity, and the same is true for $|T_n|^2$ and $|X_n - x^*|$.

(2) *L^P-convergence.* Taking expectation in (B.5.2) and summing over n, we get

$$\sup_{n\geq 1} \mathbb{E}(|T_n|^2) \leq \mathbb{E}(|T_1|^2) + s(|\beta| + b)^2. \tag{B.5.3}$$

We observe that

$$|T_{n+1}|^2 = |T_n|^2 + 2\gamma_n\langle \beta - F(X_n), T_n\rangle + 2\gamma_n\langle F(X_n) - Y_{n+1}, T_n\rangle + \gamma_n^2|\beta - Y_{n+1}|^2 \tag{B.5.4}$$

and so

$$|T_{n+1}|^2 \leq |T_n|^2 + \sum_{k=1}^{n} 2\gamma_k\langle F(X_k) - Y_{k+1}, T_k\rangle + \sum_{k=1}^{n} \gamma_k^2|Y_{k+1} - \beta|^2. \tag{B.5.5}$$

Let

$$N_n = \sum_{k=1}^{n-1} 2\gamma_k\langle F(X_k) - Y_{k+1}, T_k\rangle, \qquad n \geq 2$$

and

$$N^* = \sup_{k\geq 2} |N_k|.$$

The sequence $(N_n, \mathscr{F}_n; n \geq 2)$ is a martingale and from the Burkholder inequality it follows that for $1 \leq p < \infty$,

$$\|N^*\|_p \leq C(p) \left\| \left(\sum_{k=1}^{\infty} 4\gamma_k^2 |T_k|^2 |F(X_k) - Y_{k+1}|^2 \right)^{1/2} \right\|_p, \tag{B.5.6}$$

where $C(p)$ is the constant in the Burkholder inequality. Hence,

$$\|N^*\|_p \leq C(p) 4b\sqrt{s}\, \| \sup_{k\geq 1} |T_k| \|_p. \tag{B.5.7}$$

Put

$$T^* = \sup_{k \geq 1} |T_k|.$$

From (B.5.5), we get

$$T^{*2} \leq |T_1|^2 + N^* + s(|\beta| + b)^2 \tag{B.5.8}$$

and so it follows from (B.5.7) that

$$\|T^*\|_{2p}^2 \leq \||T_1|^2\|_p + s(|\beta| + b)^2 + C(p)\, 4b\sqrt{s}\|T^*\|_p. \tag{B.5.9}$$

By Doob's inequality (i.e., (B.5.6) with $p = 2$), we have

$$\|N^*\|_1^2 \leq \|N^*\|_2^2 \leq 64b^2 s \sup_{k \geq 1} (\mathbb{E}|T_k|^2),$$

and this last expression is finite by (B.5.3). From (B.5.8) and (B.5.9) it follows successively that

$$\|T^*\|_2 < \infty$$

and

$$\|T^*\|_p < \infty \qquad \text{for any} \quad p \in [1, \infty[;$$

an application of the Lebesgue dominated convergence theorem yields L^p-convergence.

(3) *Convergence to* x^*. It follows from (B.5.2) that

$$0 \leq \sum_{k=1}^{n} 2\gamma_k \mathbb{E}[\langle T_k, F(X_k) - \beta \rangle] \leq \mathbb{E}(|T_1|^2) + s(|\beta| + b)^2,$$

and so

$$\sum_{k=1}^{\infty} \gamma_k \mathbb{E}[\langle T_k, F(X_k) - \beta \rangle] < \infty.$$

By (H_3)

$$\sum_{k=1}^{\infty} \gamma_k \langle T_k, F(X_k) - \beta \rangle = +\infty$$

on the set $\{\delta < \lim |T_k| < 1/\delta\}$. Therefore this set is $\mathbb{P}$-negligible. □

B.5.10 Exercise
Write a program to implement the following algorithm:

$$X_1 = 2$$
$$X_{n+1} = X_n + \frac{1}{2.5 + n}\left(\frac{1}{2} - 1_{\{X_n > 10U_{n+1}\}}\right) \qquad n \geq 1,$$

where (U_n) is a Borel sequence in [0,1]. Show that the assumptions of the theorem hold with

$$F(x) = \left(\frac{x}{10} \wedge 1\right) \vee 0, \qquad \beta = \frac{1}{2}, \qquad x^* = 5.$$

B.5.11 Stopping Criteria
By running the program in Exercise B.5.10, we see that using a stochastic algorithm may be hazardous without stopping criteria. In particular, the speed of convergence may be so low that the sequence becomes in practice constant, before reaching the theoretical limit. With some additional assumptions Dvoretzky [1956] has given the following result on the rate of convergence.

Theorem B.5.12 Let a, σ and c be positive constants and assume that (again in $\mathbb{R}^d$)

$$X_{n+1} = X_n + \gamma_n(\beta - Y_{n+1}) \qquad n \geq 1,$$

with the following hypotheses

(H_1')

$$\gamma_n = \frac{ca^2}{\sigma^2 + nc^2a^2} \qquad n \geq 1$$

(H_3') Let $\mathcal{F}_1 = \sigma(X_1)$ and $\mathcal{F}_n = \sigma(X_1, \ldots, X_n, Y_2, \ldots, Y_n)$. For $n \geq 2$, there exists a Borel function $F : \mathbb{R}^d \to \mathbb{R}^d$ satisfying

$$\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = F(X_n) \qquad n \geq 1$$

and a point x^* in $\mathbb{R}^d$ such that, for all $i = 1, \ldots, d$ and for all x in $\mathbb{R}^d$, if $x_i = x_i^*$ then $F_i(x) = \beta_i$, and if $x_i \neq x_i^*$ then

$$c \leq \frac{F_i(x) - \beta_i}{x_i - x_i^*} \leq \frac{\sigma^2 + c^2a^2}{ca^2}.$$

(H_2')

$$\sup_{n \geq 1} \mathbb{E}(|F(X_n) - Y_{n+1}|^2) \leq \sigma^2.$$

Suppose moreover that

$$\mathbb{E}(|X_1 - x^*|^2) \leq a^2.$$

Then

$$\mathbb{E}(|X_n - x^*|^2) \leq \frac{\sigma^2 a^2}{\sigma^2 + (n-1)c^2a^2} \qquad n \geq 1.$$

Proof Let $T_n = X_n - x^*$ for $n \geq 1$. Then

$$\begin{aligned}\mathbb{E}(|T_{n+1}|^2) &= \mathbb{E}(|T_n|^2) + 2\gamma_n \mathbb{E}[\langle \beta - F(X_n), T_n \rangle] + \gamma_n^2 \mathbb{E}[|\beta - Y_{n+1}|^2] \\ &= \mathbb{E}(|T_n|^2) + 2\gamma_n \mathbb{E}[\langle \beta - F(X_n), T_n \rangle] \\ &\quad + \gamma_n^2 \mathbb{E}[|\beta - F(X_n)|^2] + \gamma_n^2 \mathbb{E}(|F(X_n) - Y_{n+1}|^2].\end{aligned}$$

For any x in $\mathbb{R}^d$

$$\sum_{i=1}^{n}[(x_i - x_i^*)^2 + 2\gamma_n(\beta_i - F_i(x))(x_i - x_i^*) + \gamma_n^2(F_i(x) - \beta_i)^2]$$

$$= \sum_{i=1}^{n}(x_i - x_i^*)^2\left(1 - 2\gamma_n\frac{F_i(x) - \beta_i}{x_i - x_i^*} + \gamma_n\left(\frac{F_i(x) - \beta_i}{x_i - x_i^*}\right)^2\right)$$

$$= \sum_{i=1}^{n}(x_i - x_i^*)^2\left(1 - \gamma_n\frac{F_i(x) - \beta_i}{x_i - x_i^*}\right)^2.$$

For $x_i \neq x_i^*$,

$$0 \leq 1 - \gamma_n\frac{\sigma^2 + c^2a^2}{ca^2} \leq 1 - \gamma_n\frac{F_i(x) - \beta_i}{x_i - x_i^*} \leq 1 - \gamma_n c,$$

and therefore

$$E(|T_{n+1}|^2) \leq E(|T_n|^2)(1 - \gamma_n c)^2 + \gamma_n^2\sigma^2.$$

If

$$E(|T_n|^2) \leq \frac{\sigma^2a^2}{\sigma^2 + (n-1)c^2a^2},$$

a straightforward computation proves that

$$E(|T_{n+1}|^2) \leq \frac{\sigma^2a^2}{\sigma^2 + nc^2a^2},$$

and this completes the proof. □

B.5.13 Exercise

Show that, if $0 \leq X_1 \leq 10$ and $0 < k < 10/52$, the algorithm

$$X_{n+1} = X_n + \frac{2}{5 + k.n}\left(\frac{1}{2} - 1_{\{X_n > 10.U_{n+1}\}}\right)$$

meets the requirements of Theorem B.5.12 [Hint: Show that $-(5+kn)^{-1} \leq X_{n+1} \leq 10 + (5 + kn)^{-1}$ for all $n \geq 1$, and in fact $-0.2 \leq X_n \leq 10.2$ for all $n \geq 1$.]

C LOW-DISCREPANCY SEQUENCES

C.1 Koksma–Hlawka Inequality

C.1.1 Notation

For x in $[0,1]^s$, we write $\check{x}$ the symmetrical point of x with respect to the center of the hypercube $[0,1]^s$. If $f : [0,1]^s \to \mathbb{R}$, we denote by $\check{f}$ the function $\check{f}(x) = f(\check{x})$.

Definition C.1.2 A function $f : [0,1]^s \to \mathbb{R}$ is said to have *finite variation* if there exists a bounded measure μ on the Borel sets of $[0,1]^s$, with support in $[0,1]^s \backslash \{0\}$, such that

$$\check{f}(x) = \check{f}(0) + \mu\{[[0,x]]\} \qquad \text{for all } x \text{ in } [0,1]^s.$$

This measure is unique and its mass $\|\mu\|$ is called the *variation* of f and denoted by $V(f)$.

C.1.3 Remark

A function f with finite variation is Riemann integrable. In fact, f is bounded and the discontinuity points of $\check{f}$ are contained in the countable union of hyperplanes $\{a : \mu\{x : \exists i \text{ s.t. } x_i = a_i\} \neq 0\}$.

Theorem C.1.4 (Koksma–Hlawka inequality) Let f have finite variation on $[0,1]^s$ and suppose we are given a sequence $\xi = (\xi_n)$ in $[0,1]^s$. Then, for any $N > 0$,

$$\left| \int_{[0,1]^s} f(x)\,dx - \frac{1}{N} \sum_{n=1}^{N} f(\xi_n) \right| \leq V(f) D_\infty^*(\xi, N).$$

Let $\check{\mu}$ denote the image of μ under the symmetry with respect to the center of the hypercube.

Lemma C.1.5 Let μ be a bounded measure on $[0,1]^s$, with distribution function F. We have

$$\int \pi(x)\, d\check{\mu}(x) = \int F(x)\, dx.$$

Proof of the Lemma

$$\begin{aligned}\int \pi(x)\,d\check{\mu}(x) &= \int\int 1_{[[0,x]]}(y)\,dy\,d\check{\mu}(x)\\ &= \int\int 1_{[[0,\check{x}]]}(y)\,d\mu(x)\,dy\\ &= \int\int 1_{[[0,\check{y}]]}(x)\,d\mu(x)\,dy\\ &= \int F(\check{y})\,dy.\end{aligned}$$

Proof of the Theorem Put $F = \check{f} - \check{f}(0)$. Then

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^{N} f(\xi_n) - \int f(x)\,dx &= \frac{1}{N}\sum_{n=1}^{N} \check{f}(\check{\xi}_n) - \int f(\check{x})\,dx\\ &= \frac{1}{N}\sum_{n=1}^{N} F(\check{\xi}_n) - \int F(x)\,dx\\ &= \int\left(\frac{1}{N}\sum_{n=1}^{N} 1_{\{\check{x}\le\check{\xi}_n\}} - \pi(x)\right) d\check{\mu}(x)\\ &= \int\left(\frac{1}{N}\sum_{n=1}^{N} 1_{\{\xi_n\le x\}} - \pi(x)\right) d\check{\mu}(x).\end{aligned}$$

Therefore

$$\left|\frac{1}{N}\sum_{n=1}^{N} f(\xi_n) - \int f(x)dx\right| \le D_\infty^*(\xi, N)\|\check{\mu}\| = D_\infty^*(\xi, N)\|\mu\|. \qquad \square$$

C.1.6 Remark
The same proof shows that if

$$\check{f}(x) = \check{f}(0) + \int_{[[0,x]]} h(y)\,dy$$

for some h in $L^q(dx)$, with $1 \le q < \infty$, then for $p = q/(q-1)$,

$$\left| \frac{1}{N} \sum_{n=1}^{N} f(\xi_n) - \int f(x)\, dx \right| \leq D_p^*(\xi, N) \|h\|_q.$$

C.1.7 Exercise
Let $f, g: [0,1]^s \to \mathbb{R}$, and let λ be a real number. Show that

(a) $V(f+g) \leq V(f) + V(g)$.
(b) $V(\lambda f) = |\lambda| V(f)$.
(c) $\|f\|_\infty \leq |f(1)| + V(f)$.

C.1.8 Exercise
Let $f: [0,1]^s \to \mathbb{R}$ be of class C^s. Show that

$$V(f) = \left\| \frac{\partial^s f}{\partial x_1 \cdots \partial x_s} \right\|_1.$$

C.1.9 Exercise
Define $f: [0,1]^3 \to \mathbb{R}$ by

$$f(x_1, x_2, x_3) = (x_1 + x_2 + x_3) \wedge 1.$$

Show that f does not have finite variation. [*Hint*: Show that the distribution $\partial^3 f / \partial x_1 \partial x_2 \partial x_3$ is not a measure.]

C.2 Lower Bounds of Discrepancy

Theorem C.2.1 (Roth) Let $s \geq 2$ and let ξ be any sequence in $[0,1]^s$. For every $N \geq 1$,

$$D_\infty^*(\xi, N) \geq D_2^*(\xi, N) > C(s) \frac{(\log N)^{(s-1)/2}}{N},$$

where $C(s) > 0$ is an absolute constant depending only on s.

Corollary C.2.2 Let $s \geq 1$. For any sequence ξ in $[0,1]^s$ we have

$$D_\infty^*(\xi, N) \geq C^1(s)\frac{(\log N)^{s/2}}{N}$$

for infinitely many positive integers N, where $C^1(s) > 0$ is an absolute constant depending only on s.

Proof of the Corollary Let $N \geq 1$, and consider the finite sequence $x = \{x_n; n = 1, \ldots, N\}$ in $[0, 1]^{s+1}$, where

$$x_n = \left(\xi_n, \frac{n-1}{N}\right).$$

By the above theorem, there exists $y = (y_1, \ldots, y_{s+1})$ in $[0, 1]^s \times [0, 1[$ such that

$$\left|\frac{1}{N}\sum_{n=1}^{N} 1_{[[0,y]]}(x_n) - \pi(y)\right| \geq C(s+1)\frac{(\log N)^{s/2}}{N}.$$

Now let m be the positive integer for which $(m-1)/N \leq y_{s+1} < m/N$ and put $y^{(s)} = (y_1, \ldots, y_s)$. Then

$$\begin{aligned}
&\left|\sum_{n=1}^{m} 1_{[[0,y^{(s)}]]}(\xi_n) - m\pi(y^{(s)})\right| \\
&\quad \geq \left|\sum_{n=1}^{m} 1_{[[0,y^{(s)}]]}(\xi_n) - N\pi(y)\right| - |N\pi(y) - m\pi(y^{(s)})| \\
&\quad = \left|\sum_{n=1}^{N} 1_{[[0,y]]}(x_n) - N\pi(y)\right| - |N\pi(y) - m\pi(y^{(s)})| \\
&\quad \geq C(s+1)(\log N)^{s/2} - \pi(y^{(s)})|Ny_{s+1} - m| \\
&\quad \geq C(s+1)(\log N)^{s/2} - 1
\end{aligned}$$

It follows that, for sufficiently large N, there exists m with $1 \leq m \leq N$ such that

$$mD_\infty^*(\xi, m) \geq C^1(s)(\log N)^{s/2} \geq C^1(s)(\log m)^{s/2}.$$

Letting N tend to infinity completes the proof. □

C.2.3 Remark

For logarithms to the base 2, the above statements are valid with

$$C(s) = (2^{2s+4}(s-1)^{(s-1)/2})^{-1}$$

and

$$C^1(s) = C(s+1) - \epsilon.$$

C.2.4 Exercise

For $s = 1$, show that we may not replace "infinitely many positive integers" by "for N sufficiently large" in the statement of Corollary C.2.2.

C.3 Irrational Translations on the Torus

We shall say that $\alpha = (\alpha_1, \ldots, \alpha_s)$ in $\mathbb{R}^s$ is *irrational* if $1, \alpha_1, \ldots, \alpha_s$ are linearly independent over the rational numbers. Let α be irrational and, for x in $\mathbb{R}^s$, let $\xi_n = \{x + n\alpha\}$ for $n \geq 1$. If $m = (m_1, \ldots, m_d) \neq 0$, and $N \geq 1$, we have

$$\frac{1}{N}\sum_{n=1}^{N} e^{2i\pi\langle m,\xi_n\rangle} = \frac{1}{N}\, e^{2i\pi\langle m,x\rangle}\, \frac{e^{2i\pi\langle m,\alpha\rangle(N+1)} - 1}{e^{2i\pi\langle m,\alpha\rangle} - 1}.$$

By the Weyl criterion, the sequence $\xi = (\xi_n)_{n\geq 1}$ is u.d. in $[0,1]^s$ as in Example A.5.4.1. We recall the following definition.

Definition C.3.1 Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space. A measurable transformation $\theta\colon \Omega \to \Omega$ is called measure-preserving if $\mathbb{P}(\theta^{-1}(A)) = \mathbb{P}(A)$ holds for all A in $\mathscr{A}$. A measure-preserving transformation θ of Ω is called ergodic if for every set A in $\mathscr{A}$ with $\theta^{-1}(A) = A$ $\mathbb{P}$-a.s., we have $\mathbb{P}(A) = 0$ or 1.

Lemma C.3.2 Let θ be a measure-preserving transformation and let $U\colon f \mapsto f \circ \theta$ be the associated linear isometric operator on $L^2(\Omega, \mathscr{A}, \mathbb{P})$. Then θ is ergodic iff the equation $Uf = f$, $f \in L^2$, has only constant solutions.

Proof Exercise.

We shall identify $[0,1[^s$ with the s-dimensional torus $\mathbf{T}_s = \mathbb{R}^s/\mathbf{Z}^s$ endowed with the finest topology for which the canonical projection $\mathbb{R}^s \to \mathbf{T}_s$ is continuous. This topology is compatible with the metric

$$d((x_1,\ldots,x_s),(y_1,\ldots,y_s)) = \sup_{i=1,\ldots,s} \{\min(|x_i - y_i|, 1 - |x_i - y_i|)\}$$

which is translation invariant.

Theorem C.3.3 The translation $\theta_\alpha: x \mapsto \{x + \alpha\}$ in $\mathbf{T}_s$ is ergodic iff α is irrational.

Proof For m in $\mathbf{Z}^s$ put $\psi_m(x) = e^{2i\pi\langle m,x\rangle}$. The functions $(\psi_m, m \in \mathbf{Z}^s)$ set up an orthonormal basis in $L^2(\mathbf{T}_s, \lambda_s)$. For every f in $L^2(\mathbf{T}_s, \lambda_s)$,

$$\|f \circ \theta_\alpha\|_2 = \|f\|_2.$$

It follows that θ_α is measure-preserving and the associated operator U_α satisfies

$$U_\alpha \psi_m = \psi_m \circ \theta_\alpha = e^{2i\pi\langle m,\alpha\rangle}\psi_m.$$

It follows from the lemma that a necessary condition for θ_α to be ergodic is that

$$e^{2i\pi\langle m,\alpha\rangle} \neq 1 \qquad \text{for all} \quad m \neq 0,$$

which is equivalent to the irrationality of α. Conversely, let us assume that α is irrational, and let f in $L^2(\mathbf{T}_s, \lambda_s)$ be represented by its Fourier series

$$f = \sum_{m \in \mathbf{Z}^s} \hat{f}(m)\psi_m.$$

If f is invariant, then

$$0 = U_\alpha f - f = \sum_{m \in \mathbf{Z}^s} (e^{2i\pi\langle m,\alpha\rangle} - 1)\hat{f}(m)\psi_m.$$

Therefore $\hat{f}(m) = 0$ for all $m \neq 0$, and f is constant. □

It follows from the pointwise ergodic theorem that for every f in $L^1(\mathbf{T}_s, \lambda_s)$,

$$\frac{1}{N}\sum_{n=1}^{N} f(\{x + n\alpha\}) \to_{N\uparrow\infty} \int f(y)\,dy$$

for λ_s-almost every x in $\mathbf{T}_s$. If in addition f is Riemann integrable, we know that convergence holds everywhere in $\mathbf{T}_s$. Moreover, the following statement is true.

Proposition C.3.4 Let α be irrational in $\mathbb{R}^s$.

(i) The only probability measure μ on $\mathbf{T}_s$ such that $(\theta_\alpha)_*\mu = \mu$ is the Lebesgue measure λ_s.

(ii) For all f in $\mathscr{C}(\mathbf{T}_s)$,

$$\lim_{N\uparrow\infty} \sup_x \left| \frac{1}{N}\sum_{n=1}^{N} f \circ \theta_\alpha^n(x) - \int f(y)\,dy \right| = 0.$$

Proof (i) Let f be in $\mathscr{C}(\mathbf{T}_s)$ and let μ be a probability measure on $\mathbf{T}_s$ such that $(\theta_\alpha)_*\mu = \mu$. Then

$$\left\langle \mu, \frac{1}{N}\sum_{n=1}^{N} f \circ \theta_\alpha^n \right\rangle = \mu(f).$$

It follows from the pointwise convergence of $(1/N)\sum_{n=1}^{N} f \circ \theta_\alpha^n$ to $\int f(y)\,dy$ and the Dominated Convergence Theorem that

$$\int f(y)dy = \mu(f),$$

and this proves the first statement.

(ii) Set

$$\varphi_N = \frac{1}{N}\sum_{n=1}^{N} f \circ \theta_\alpha^n.$$

It follows from the uniform continuity of f on the compact Hausdorff space $\mathbf{T}_s$ that, for every $\epsilon > 0$, there exists $h > 0$ such that

$$d(x, y) < h \Rightarrow |f(x) - f(y)| < \epsilon$$

and therefore

$$|\varphi_N(x) - \varphi_N(y)| \leq \frac{1}{N} \sum_{n=1}^{N} |f \circ \theta_\alpha^n(x) - f \circ \theta_\alpha^n(y)| < \epsilon.$$

The uniform convergence of $(\varphi_N)$is now a consequence of pointwise convergence and of Ascoli's theorem. □

C.3.5 Remark

It is worth noting that statements (i) and (ii) are equivalent for any compact metric space E endowed with its Borel σ-field $\mathscr{E}$, a probability measure μ and a measure-preserving transformation θ [Cornfeld, Fomin, and Sinai, 1982].

C.3.6 Exercise

There is another way to prove part (ii) of Proposition C.3.4. If f is a trigonometric polynomial with no constant term, show that the sum $\sum_{n=1}^{N} f \circ \theta_\alpha^n$ remains bounded and use the Stone–Weierstrass theorem to complete the proof.

C.3.7 Remark

For any sequence $\xi = (\xi_n)_{n \geq 1}$ of real numbers in [0, 1], set

$$\varphi(\xi) = \sup_N \frac{N}{\log N} D_\infty^*(\xi, N).$$

Then [Dupain and Sós, 1980]

$$\varphi((\{n\sqrt{2}\})) = (4 \log (1 + \sqrt{2}))^{-1}$$

and this is the minimum value among all the numbers $\varphi((\{n\sqrt{\alpha}\}))$.

C.3.8 Examples

(1) [Hecke, 1922] for $s = 1$, [W.M. Schmidt, 1970]. Let $\alpha = (\alpha_1, \ldots, \alpha_s)$ be an irrational such that each α_i is an algebraic real number, and let

$\xi = (\{n\alpha\})$. Then, for every $\epsilon > 0$, there exists $C(\alpha, \epsilon) > 0$ such that

$$D_\infty^*(\xi, N) \leq C(\alpha, \epsilon)\frac{1}{N^{1-\epsilon}}.$$

(2) [A. Baker, 1965]. Let $\beta = (\beta_1, \ldots, \beta_s)$ with $\beta_i = e^{r_i}$, where the r_i are distinct non-zero rational numbers. If $\xi = (\{n\beta\})$, then for any $\epsilon > 0$ there exists a number $C(\beta, \epsilon)$ such that

$$D_\infty^*(\xi, N) \leq C(\beta, \epsilon)\frac{1}{N^{1-\epsilon}}.$$

C.3.9 Points of the Torus with Rational Coordinates

It is possible to show, using the property of unique factorization in the ring $Z[i]$ of Gaussian integers, that the algorithm

$$\begin{aligned}
x_{1,n+1} &= \tfrac{3}{5}x_{1,n} - \tfrac{4}{5}y_{1,n} \\
y_{1,n+1} &= \tfrac{4}{5}x_{1,n} + \tfrac{3}{5}y_{1,n} \\
x_{2,n+1} &= \tfrac{5}{13}x_{2,n} - \tfrac{12}{13}y_{2,n} \\
y_{2,n+1} &= \tfrac{12}{13}x_{2,n} + \tfrac{5}{13}y_{2,n},
\end{aligned}$$

starting from a point $((x_{1,0}, y_{1,0}), (x_{2,0}, y_{2,0}))$ of the torus $\mathbf{T}_2$, defines an u.d. sequence.

More generally, the sequence $\zeta_n = (\zeta_1^n, \ldots, \zeta_s^n)$, where the irreducible form

$$\zeta_j = \frac{p_j}{r_j} + \frac{qj}{r_j}$$

involves in the decomposition of r_j a prime number that does not appear in that of $r_1, \cdots, r_{j-1}$ [Bouleau, (11)].

C.4 Other Usual Sequences with Low Discrepancy

C.4.1 Halton Sequence

Let $p \geq 2$ be an integer. If the p-adic expansion of n in $\mathbf{N}^*$ is

$$n = a_0 + a_1 p + \cdots + a_k p^k,$$

where $a_i \in \{0, \ldots, p-1\}$ for $i = 0, \ldots, k$, then $\varphi_p(n)$ is defined by

$$\varphi_p(n) = \frac{a_0}{p} + \frac{a_1}{p^2} + \cdots \frac{a_k}{p^{k+1}}.$$

A Halton sequence $\xi = (\xi_n)$ in $[0,1]^s$ is defined by

$$\xi_n = (\varphi_{p_1}(n), \ldots, \varphi_{p_s}(n)), \qquad n \geq 1,$$

where $p_1, \ldots, p_s$ are integers that are greater than 1 and pairwise relatively prime. The following discrepancy estimate holds [Meijer, 1968].

$$D_\infty^*(\xi, N) \leq \frac{1}{N}\left[2\sum_{i=1}^{s} p_i + (2\log N)^s \prod_{i=1}^{s} \frac{p_i - 1}{\log p_i}\right], \qquad n \geq 1.$$

This means that

$$D_\infty^*(\xi, N) = O(N^{-1}(\log N)^s).$$

C.4.2 Hammersley Sequence

The Hammersley sequence of order N in $[0,1]^s (s \geq 2)$ is defined by

$$\xi_n = \left(\frac{n}{N}, \varphi_{r_1}(n), \ldots, \varphi_{r_{s-1}}(n)\right), \qquad 1 \leq n \leq N,$$

where $r_1, \ldots, r_{s-1}$ are the first $s-1$ primes. In this case

$$D_\infty^*(\xi, N) \leq \frac{1}{N} \prod_{i=1}^{s-1} \frac{r_i \log N r_i}{\log r_i} \qquad \text{if} \quad N > r_{s-1}.$$

C.4.3 Lapeyre–Pagès Sequences

Let $p \geq 2$ be an integer. For real number x in $[0,1[$, the p-adic expansion

$$x = \sum_{k=0}^{\infty} \frac{x_k}{p^{k+1}}$$

is said to be proper if the set $\{k : x_k \neq p-1\}$ is infinite. If

$$x = \sum_{k=0}^{\infty} \frac{x_k}{p^{k+1}} \qquad \text{and} \qquad y = \sum_{k=0}^{\infty} \frac{y_k}{p^{k+1}}$$

are the proper p-adic expansions of x and y in $[0,1[$, let $x \oplus_p y$ be the number z in $[0, 1]$ with p-adic expansion

$$z = \sum_{k=0}^{\infty} \frac{z_k}{p^{k+1}}$$

where

$$\begin{aligned} z_0 &= x_0 + y_0 && \bmod p \\ z_n &= x_n + y_n && \bmod p \quad \text{if} \quad x_{n-1} + y_{n-1} < p \\ &= 1 + x_n + y_n && \bmod p \quad \text{if} \quad x_{n-1} + y_{n-1} \geq p. \end{aligned}$$

Now suppose that $p_1, \ldots, p_s$ are greater than 1 and pairwise relatively prime. Let $y_1, \ldots, y_s$ be numbers in $[0, 1[$ such that, for $i = 1, \ldots, s$, y_i has a finite p_i-adic expansion and lies in $[1/p_i, 1[$. For x in $[0, 1[$, put

$$T_{y_i}(x) = y_i \oplus_{p_i} x.$$

Finally, for α in $\mathbf{T}_s$, consider the sequence $\xi = (\xi_n)$ defined by

$$\xi_n = (T_{y_1}^{n-1}(\alpha_1), \ldots, T_{y_s}^{n-1}(\alpha_s)).$$

The following estimate holds:

$$D_\infty^*(\xi, N) \leq \frac{1}{N}\left[1 + \prod_{i=1}^{s} (p_i - 1)\, \frac{\log p_i N}{\log p_i}\right].$$

Note that this estimate does not depend on α.

C.4.4 Faure Sequences

Let $s \geq 1$ and let $p \geq s$ be a prime number. Let Δ_p be the set of p-adic rationals in $[0,1[$ (i.e., those with finite p-adic expansion). Define $C_p: \Delta_p \to \Delta_p$ for

$$x = \frac{a_0}{p} + \cdots + \frac{a_k}{p^{k+1}}$$

by

$$C_p(x) = \frac{b_0}{p} + \cdots + \frac{b_k}{p^{k+1}},$$

where

$$b_j = \sum_{i=j}^{k} \frac{i!}{j!(i-j)!}\, a_i \qquad \mod p.$$

Set

$$\begin{aligned} C_i^j &= 0 && \text{if} \quad j > i \geq 0 \\ &= \frac{i!}{j!(i-j)!} && \text{if} \quad i \geq j \geq 0. \end{aligned}$$

Lemma C.4.5 For $n \geq 1$, the infinite matrix $C = (C_i^j)$ satisfies

$$(C^n)_{ij} = n^{i-j} C_i^j.$$

Proof We proceed by induction. The statement holds for $n = 1$. Suppose that it holds for some $n \geq 1$. Then

$$\begin{aligned} (C^{n+1})_{ij} &= \sum_{k=j}^{i} (C^n)_{ik} C_k^j = \sum_{k=j}^{i} n^{i-k} C_i^k C_k^j \\ &= \sum_{k=j}^{i} n^{i-k} C_i^j C_{i-j}^{k-j} = C_i^j \sum_{l=0}^{i-j} C_{i-j}^l n^{i-j-l} \\ &= C_i^j (n+1)^{i-j}, \end{aligned}$$

so it holds for $n + 1$. □

Thus, it turns out that the map C_p is injective and satisfies

$$(C_p)^p = Id \qquad (C_p)^{-1} = (C_p)^{p-1}.$$

C.4.6 Exercise

Is the mapping C_p extendable to a Riemann integrable mapping from [0, 1] into [0,1]?

Proposition C.4.7 If $s \geq 2$ and $p \geq 3$ is a prime such that $p \geq s$, then the sequence

$$\xi_n = (\varphi_p(n), C_p(\varphi_p(n)), \ldots, C_p^{s-1}(\varphi_p(n)))$$

satisfies

$$D_\infty^*(\xi, N) \leq \frac{1}{N}\left[\frac{1}{s!}\left(\frac{p-1}{2\log p}\right)^s (\log N)^s + O((\log N)^{s-1})\right].$$

We refer to H. Faure, [1982] for a complete proof of this proposition. Here we only derive an interesting geometric property of Faure sequences.

Proposition C.4.8 Let $m \in \mathbf{N}$ and $l \in \mathbf{N}^*$. For any choice of $r_1, \ldots, r_s$ and $u_1, \ldots, u_s$ in $\mathbf{N}$ with $r_1 + \cdots + r_s = m$ and $u_k < p^{r_k}$, there is exactly one element of the sequence $\xi_{lp^m}, \xi_{lp^m+1}, \ldots, \xi_{lp^m+p^m-1}$ in the s-dimensional interval $I = \prod_{k=1}^s [u_k/p^{r_k}, (u_k+1)/p^{r_k}[$.

Proof For $n \geq 1$, let

$$n = \sum_{i=0}^{\infty} a_i(n) p^i$$

be the p-adic expansion of n. Note first that, for $i \geq m$, each $a_i(n)$ is uniquely determined by the condition $lp^m \leq n < lp^m + p^m$. For $k = 1, \ldots, s$, we have

$$\xi_{n,k} = C_p^{k-1}(\varphi_p(n)) = \sum_{j=0}^{\infty} \frac{y_j^{(k)}(n)}{p^{j+1}},$$

where

$$y_j^{(k)}(n) = \sum_{i=j}^{\infty} C_i^j (k-1)^{i-1} a_i(n) \qquad \mod p.$$

Existence and uniqueness of $y_j^{(k)}(n)$ for $0 \le j \le r_k - 1$ follow from the condition

$$\frac{u_k}{p^{r_k}} \le \xi_{n,k} < \frac{u_k + 1}{p^{r_k}}.$$

Therefore the vector $(a_i(n), 0 \le i \le m-1)$ is a solution of the following system of equations in $\mathbf{Z}/p\mathbf{Z}$.

$$\begin{cases} \displaystyle\sum_{i=j_k}^{\infty} C_i^{j_k} (k-1)^{i-j_k} a_i(n) = y_{j_k}^{(k)}(n) \qquad \mod p \\ k = 1, \ldots, s \\ j_k = 0, \ldots, r_k - 1 \end{cases}$$

This is a system of $\sum_{k=1}^{s} r_k = m$ equations with m unknown quantities. Its determinant is a kind of Van der Monde determinant and its value is

$$\prod_{1 \le h < k \le s} (k-h)^{r_k r_h} \neq 0.$$

Accordingly, the problem of finding n such that ξ_n belongs to I and

$$lp^m \le n < (l+1)p^m$$

has exactly one solution. □

C.4.9 Approximate Sequences

In practice we are often forced to use sequences (y_n) to approximate sequences (ξ_n) which have suitable asymptotic behavior, but do not have rational components.

Lemma C.4.10 Let $\xi = (\xi_n)$ and $\eta = (\eta_n)$ be two sequences in $[0,1]^s$ such that

$$\sup_{i=1,\dots,s} |\xi_{n,i} - \eta_{n,i}| \le h(n).$$

Then, for every p in $[1, \infty[$ and every $N \ge 1$,

$$|D_p^*(\xi, N) - D_p^*(\eta, N)| \le \frac{s^{1/p}}{N} \sum_{n=1}^{N} (h(n))^{1/p}.$$

Proof

$$\left| \frac{1}{N} \sum_{n=1}^{N} 1_{[[0,x]]}(\xi_n) - \pi(x) \right| \le \left| \frac{1}{N} \sum_{n=1}^{N} 1_{[[0,x]]}(\eta_n) - \pi(x) \right| + \frac{1}{N} \sum_{n=1}^{N} |1_{[[0,x]]}(\xi_n) - 1_{[[0,x]]}(\eta_n)|$$

Take now the L^p-norm to obtain

$$D_p^*(\xi, N) \le D_p^*(\eta, N) + \frac{1}{N} \sum_{n=1}^{N} (sh(n))^{1/p}.$$

It is clear that if $h(n) \to 0$ as n tends to infinity, then the sequences ξ and η are simultaneously u.d. We have already seen in Section C.3 a metric on the torus $\mathbf{T}_s$ for which the canonical projection $x \in \mathbb{R}^s \to \{x\} \in \mathbf{T}_s$ is continuous. Now suppose that $\xi = (\xi_n)$ and $\eta = (\eta_n)$ are sequences in $[0, 1]^s$ such that $d(\xi_n, \eta_n) \le h(n)$ for every n. We write $r(n)$ for the number of indices $k \le n$ such that the distance $d(\xi_k, B)$ from ξ_k to the edge of the cube $[0, 1]^s$ is no more than $h(k)$. Then

$$|D_p^*(\xi, N) - D_p^*(\eta, N)| \le \frac{s^{1/p}}{N} \sum_{k=1}^{N} h^{1/p}(k) + \frac{r(N)}{N}.$$

If ξ is u.d. and $h(n) \to 0$ as n tends to infinity, then $r(N)/N \to 0$ as $N \uparrow \infty$ and the sequence η is also u.d.

C.4.11 Exercises

(1) Let $p_1, \ldots, p_s$ be the first s prime numbers and, for $k = 1, \ldots, s$, let the sequences x^k be defined by

$$x_0^k = [\sqrt{p_k}]$$
$$x_{n+1}^k = \frac{1}{2}\left(\frac{p_k}{x_n^k} + x_n^k\right).$$

Show that the sequence $(\{nx_n^1\}, \ldots, \{nx_n^s\})$ is u.d. in $[0, 1]^s$ (it is an approximate sequence of $(\{n\sqrt{p_1}\}, \ldots, \{n\sqrt{p_s}\})$).

(2) Let $\alpha = (e^{r_1}, \ldots, e^{r_s})$, where the r_i are distinct rational numbers. By a result of A. Baker (cf., A.5.4.1)) the sequence $\{n\alpha\} = (\{ne^{r_1}\}, \ldots, \{ne^{r_s}\})$ is u.d. in $[0, 1]^s$. Show that the sequence $(\{na_n^1\}, \ldots, \{na_n^s\})$ defined by

$$a_n^k = a_{n-1}^k + r_n^k$$
$$r_n^k = r_{n-1}^k \frac{r_k}{n}$$

with $a_0^k = r_0^k = 1$, is u.d. in $[0, 1]^s$.

To be useful in practice, the sequences x_n^k and a_n^k in the previous exercises must be computed with increasing accuracy (like $\log_2 n$) by representing a real number by an array of digits.

C.5 Low-Discrepancy Sequences and Simulation

(a) Suppose that X is a uniformly distributed random variable on $[0, 1]^s$ and f is a Riemann-integrable function on $[0, 1]^s$. Instead of computing $\mathbb{E}(f(X))$ by the law of large numbers, that is,

$$\mathbb{E}(f(X)) = \lim_{N \uparrow \infty} \frac{1}{N} \sum_{n=1}^{N} f(U_n) \qquad \text{a.s.}$$

where (U_n) is a Borel sequence in $[0, 1]^s$, we may replace (U_n) with a uniformly distributed sequence (ξ_n) in $[0, 1]^s$.

C.5.1 Remarks

(1) As $U_n = (U_{n,1}, \ldots, U_{n,s})$, this amounts to replacing the $(ns+j)$th call to the random function (i.e., the pseudo-random numbers generator) with

$\xi_{n,j} (1 \le j \le s)$. It would be wrong to take a u.d. sequence in $[0, 1]$ and use it instead of the sequence of calls to the random function.

(2) The advantage of this method is that it goes faster, provided that the sequence (ξ_n) is suitable and f is not too irregular; furthermore, there is a deterministic stopping criterion, the Koksma–Hlawka inequality.

C.5.2 Exercise

For the Van der Corput sequence $x_n = \varphi_2(n)$, show that

$$\lim_{N \uparrow \infty} \frac{1}{N} \sum_{n=1}^{N} \left(x_{2n} - \frac{1}{2} \right) \left(x_{2n+1} - \frac{1}{2} \right) = -\frac{1}{24},$$

so this sequence is not pseudo-random. [Hint: express x_{2n} and x_{2n+1} as functions of x_n.]

(b) Another typical example of the efficiency of low discrepancy sequences is given by simulating a real random variable X by a rejection method. Let (U_n) be a Borel sequence in $[0, 1]^s$, let A be a Borel set in $[0, 1]^s$ with negligible boundary and positive Lebesgue measure, and let $T_n = \inf\{n \ge 1 : U_n \in A\}$. Suppose that there is a Riemann integrable function f such that the distribution of the random variable X is given by $f(U_T)$. We have seen the distribution of U_T is uniform in A and so

$$\mathbb{E}X = \frac{1}{|A|} \int_A f(x)\, dx.$$

In the classical Monte Carlo method, the value of $\mathbb{E}X$ is evaluated by

$$\mathbb{E}X = \lim_{N \uparrow \infty} \frac{\sum_{n=1}^{N} f(U_n) 1_A(U_n)}{\sum_{n=1}^{N} 1_A(U_n)} \quad \text{a.s.}$$

If $\xi = (\xi_n)$ is u.d. in $[0, 1]^s$, however, we also have (exercise)

$$\mathbb{E}X = \lim_{N\uparrow\infty} \frac{\sum_{n=1}^{N} f(\xi_n) 1_A(\xi_n)}{\sum_{n=1}^{N} 1_A(\xi_n)};$$

that is, the Borel sequence in $[0, 1]^s$ can be replaced by successive calls to a low discrepancy sequence.

C.5.3 Comparison Between the Monte Carlo Method and Quasi-Monte Carlo Methods

Difference of Nature Between the Two Approaches The use of pseudo-random numbers for computing expectations is quite a common technique. It works well for computations with ordinary accuracy, which is often all that is wanted when engineers use probabilistic models.

There are, however, several points that are worth keeping in mind.

(1) No explicit sequence of digits has been proved to have a random character. There are deep theoretical reasons for this [Martin-Löf, 1966; Dellacherie, 1978; Knuth, 1981].

(2) We can ask not that the sequence satisfy every statement in set theory which is true almost surely, but only that it should satisfy the main ones: normal character in the sense of Borel, the χ-square law, and these properties for images of the sequence under a simple family of functions. Even in this case, there are very few sequences for which these properties have been proved (Champernowne sequence or improvements of it [Rauzy, 1981, Dumont and Thomas, 1986], etc.), and in practice they are far from being good pseudo-random sequences.

(3) In practice, pseudo-random sequences are obtained in the following way. Families of generators are examined by statistical tests and the sequences that behave badly are eliminated. There are a great number of tests [Knuth, 1981; Niederreiter, 1978; Ripley 1987; Fushimi, 1988; Marsaglia and Zaman, 1990; Anderson, 1990; Altman, 1988; Marsaglia, 1985; Ripley, 1990; Flajolet, Kirschenhofer, and Tichy, 1988]. Thus one gets sequences which can be considered as convenient. Most of them are finite (i.e., periodic) sequences but it is also possible to build infinite sequences (cf., Exercise B.1.3).

The Monte Carlo method works by using these convenient sequences for computations in the framework of convergence theorems, which are essentially just new statistical tests that the sequence has not yet passed. If the sequence has been used for a long time, then it is reasonable to

think that if it had anomalies they would already be detected. Nevertheless, the need for longer and longer finite sequences and the imagination of mathematicians yield many new finite or infinite sequences; for instance, more and more algorithms (discrete dynamical systems) are known to have chaotic behavior.

There is therefore a bet involved in applying the Monte Carlo method, which must in consequence be used carefully. The equidistributed sequences used in quasi-Monte Carlo methods are quite different in nature, in that the equidistribution property is often proved by a demonstration that yields an explicit bound of their discrepancy. In other words, only one property (equidistribution) of the sequence is used in the quasi-Monte Carlo method, and this property is proved to hold. On the other hand, the Monte Carlo method is reliable so long as the mathematical model to be computed has no relationship with the dynamics of the sequence being used. Unfortunately, no rigorous analysis of this "independence" is yet available and it is therefore necessary to check that the results do not depend on the family of generators and to check that they are continuous with respect to the parameters of the model (when such a continuity holds).

> Si haut que l'on remonte dans l'histoire des sciences, on voit la certitude que donnent les mathématiques leur valoir un rang éminent dans l'esprit des hommes.[3] [Henri Lebesgue, 1938]
>
> In mathematics we are lucky to have proofs at all! Other subjects are not so fortunate. [Laurence C. Young, 1981]

Quantitative Comparison of the Speed of Two Methods This sort of comparison was made in Sarkar and Prasad [1987] and Pagès and Xiao [1991]. These works show that, for all currently known families of low-discrepancy sequences on $[0,1]^s$, the number of drawings required for these sequences to equal (and then surpass) the accuracy obtained by pseudo-random sequences increases rather quickly with the dimension s. Thus the use of low-discrepancy sequences is in practice limited to low dimensional problems. We extract from the study quoted in Pagès and Xiao [1991] the following results:

In Figure 2.2, the expression "k times better than random" does not mean that the sequence is compared with a random sequence, which would give a random result, but that it is compared with the asymptotic speed of random sequences given by the Chung–Kiefer theorem (Theorem B.4.2).

[3]However far we go back through the history of science, we see how the certainty given by mathematics affords it a high rank in the minds of men.

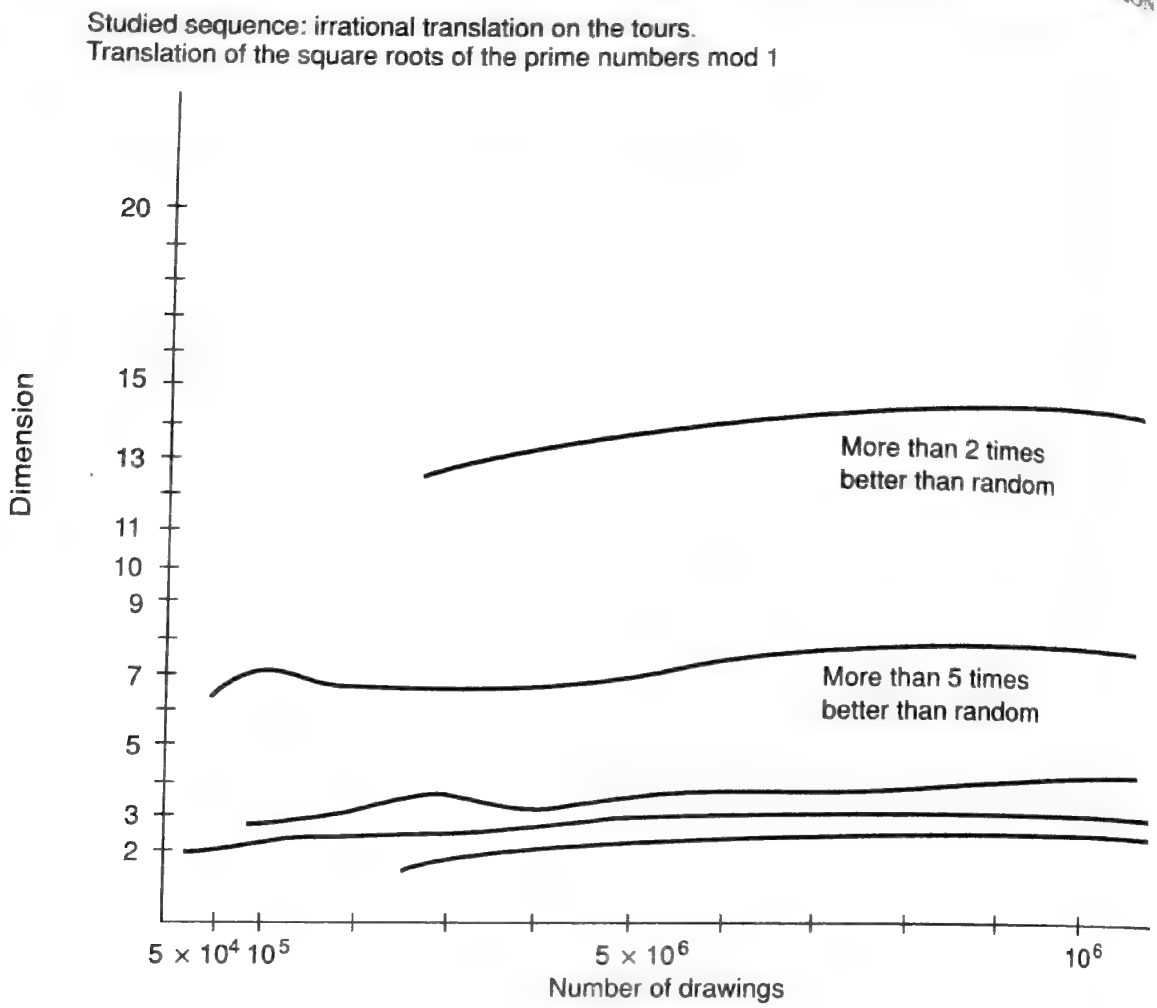


Figure 2.2

C.6 Complement: Boreloid Sequences

We have seen in Section C.5 that random sequences converge more slowly for integration than low-discrepancy sequences, at least when the dimension s is small. Nevertheless, they can be used for very irregular functions since, by the strong law of large numbers,

$$\forall f \in L^1([0,1]^s, dx) \qquad \mathbb{E}f = \lim_{N \uparrow \infty} \frac{1}{N} \sum_{n=1}^{N} f(U_n) \qquad \text{a.s.}$$

for every Borel sequence $(U_n)_{n \geq 1}$ on $[0,1]^s$.

Thus it is natural to ask whether there exist random sequences $X = (X_n(\omega))$, which, like Borel sequences, can be used to integrate irregular functions and, as the low discrepancy sequences, have the property of running fast.

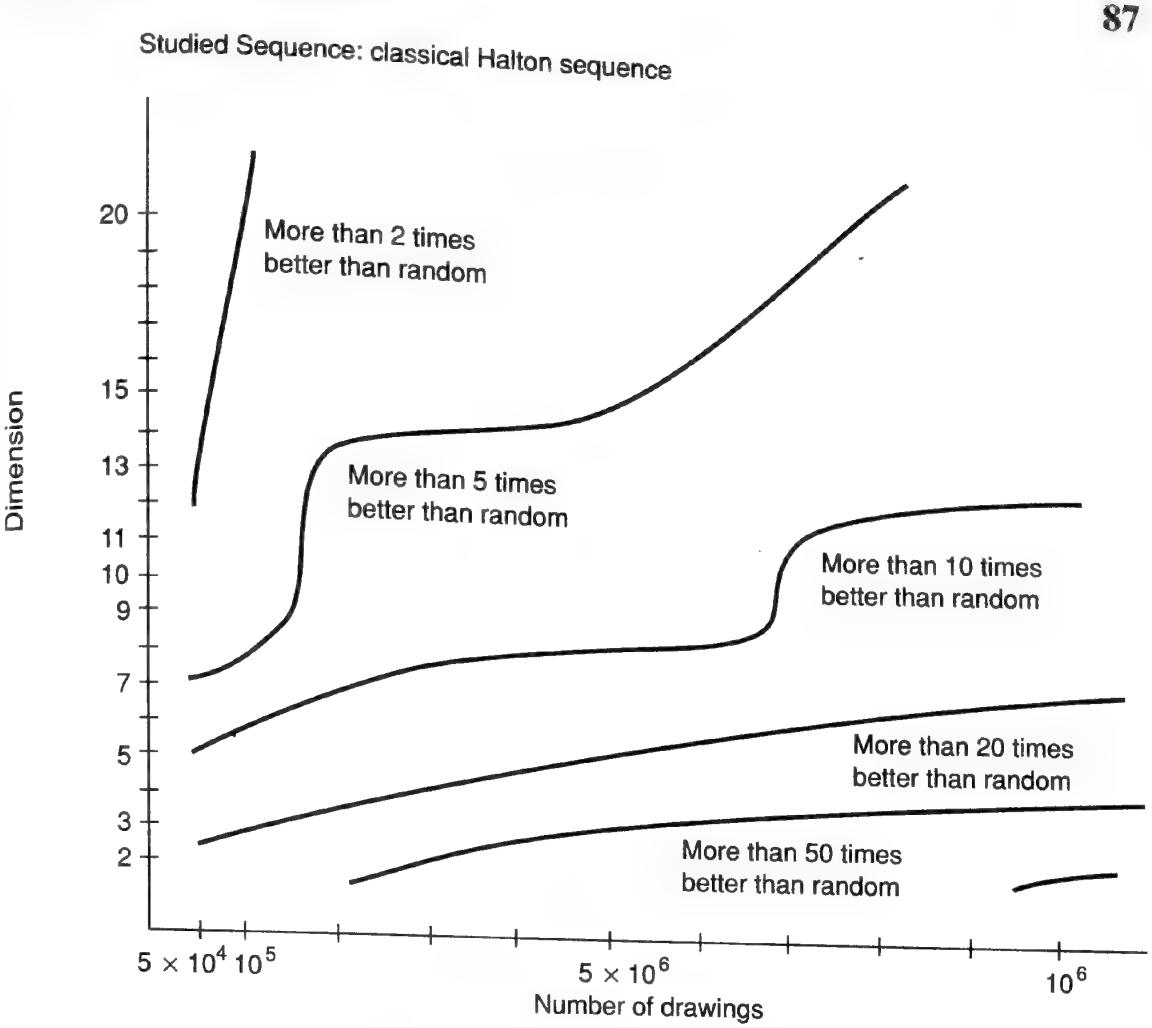


Figure 2.2 (*Continued*)

We emphasize the fact that the study presented here has a theoretical character. (It will become apparent in Chapters 3 and 5 that integrating very irregular Borel functions poses practical problems even with good pseudo-random sequences.)

Definition C.6.1 Let μ be a probability measure on $([0,1]^s, \mathcal{B}[0,1]^s)$.

(i) A sequence of random variables (X_n) with values in $[0,1]$ is said to be *almost surely* μ-*distributed*, if the sequence $(X_n(\omega))_{n\geq 1}$ is μ-distributed on $[0,1]^s$ with probability one.

(ii) A sequence of random variables (X_n) with values in $[0,1]^s$ is said to be *Boreloid* μ-*distributed* if, for every bounded Borel function f from $[0,1]^s$ into $\mathbb{R}$,

$$(*) \qquad \frac{1}{N}\sum_{n=1}^{N} f(X_n) \to_{N\uparrow\infty} \int f\, d\mu \qquad \text{a.s.}$$

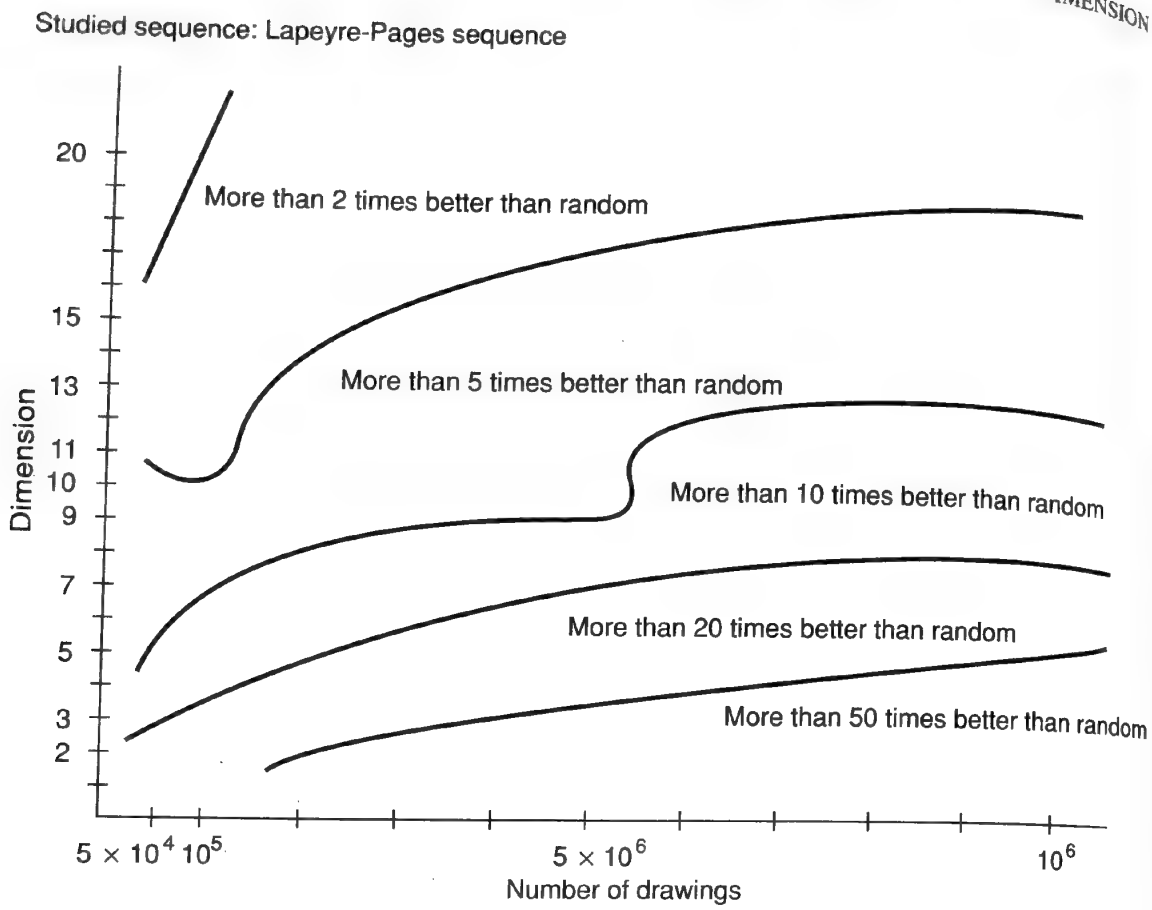


Figure 2.2 (*Continued*)

Using the same argument that derives the fundamental theorem of statistics (also called Glivenko–Cantelli theorem; cf. Theorem B.2.1) from the law of large numbers, we see that if a sequence (X_n) is Boreloid μ-distributed then it is almost surely μ-distributed. Indeed, for such a sequence we have

$$\lim_N \frac{1}{N} \sum_{n=1}^{N} f(X_n) = \int f \, d\mu \qquad \text{a.s.}$$

for every f in a countable dense subset D of $\mathscr{C}([0,1]^s)$ and so

$$\text{a.s.} \qquad \lim_N \frac{1}{N} \sum_{n=1}^{N} f(X_n) = \int f \, d\mu \qquad \forall f \in D.$$

This easily implies

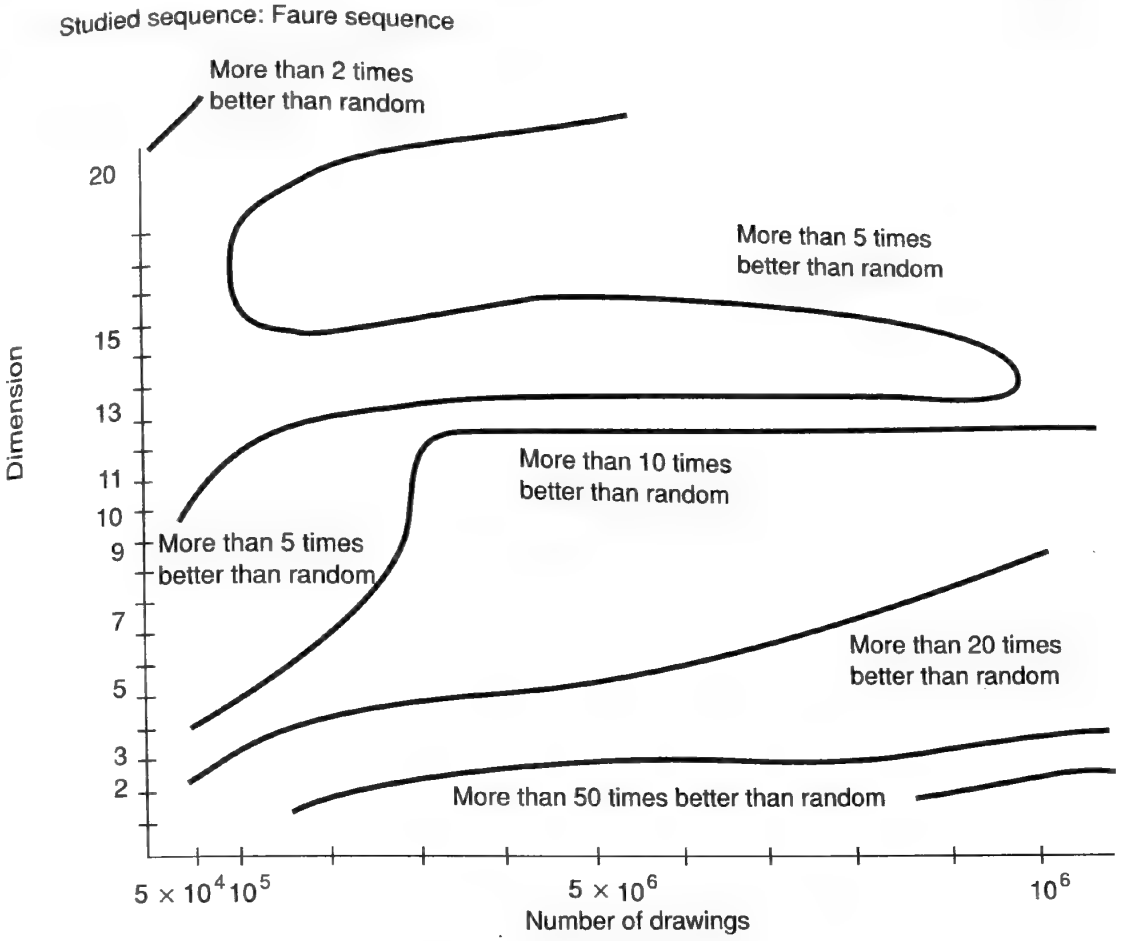


Figure 2.2 (*Continued*)

$$\text{a.s.} \qquad \frac{1}{N}\sum_{n=1}^{N}\delta_{X_n} \to \mu \text{ narrowly.}$$

Proposition C.6.2 Let (X_n) be a sequence of random variables with values in $[0,1]^s$. Suppose that for every open set A in $[0,1]^s$

$$(**) \qquad \frac{1}{N}\sum_{n=1}^{N} 1_A(X_n) \to \mu(A) \qquad \text{a.s.}$$

Then (X_n) is Boreloid μ-distributed.

For the proof we use the following lemma.

Lemma C.6.3 Property $(**)$ for open sets implies property $(*)$ for all continuous f.

Proof of Lemma Let f be continuous. We may assume $0 < f < 1$.

Then

$$f^{-1}\left(\left[\frac{k}{n}, \frac{k+1}{n}\right[\right) = G_{k+1}\backslash G_k,$$

where the G_k are increasing open sets. We uniformly approximate f up to ϵ by

$$f_K = \sum_{k=1}^{K} \alpha_k 1_{A_k},$$

where $A_k = G_{k+1}\backslash G_k$. Then f_K satisfies $((1/N)\sum_{n=1}^{N} f_K(X_n) \to \int f_K d\mu$ a.s.); hence

$$\overline{\lim} \frac{1}{N} \sum_{n=1}^{N} f(X_n) \leq \int f \, d\mu + 2\epsilon$$

$$\underline{\lim} \frac{1}{N} \sum_{n=1}^{N} f(X_n) \geq \int f \, d\mu - 2\epsilon,$$

which gives the lemma. □

Proof of Proposition To prove the proposition, we recall the Lusin property for a bounded Borel function f: For every sequence (ϵ_k) decreasing to zero, there exists a sequence of open sets G_k and a sequence of continuous functions f_k such that $\forall k$,

$$\mu(G_k) \leq \epsilon_k$$
$$f = f_k \text{ outside } G_k$$
$$\|f_k\|_\infty \leq \|f\|_\infty$$

[Rudin, 1966]. Now

$$\frac{1}{N} \sum_{n=1}^{N} f(X_n) = \frac{1}{N} \sum_{n=1}^{N} f_k(X_n) + \frac{1}{N} \sum_{n=1}^{N} (f - f_k)(X_n).$$

By the lemma, the following inequalities are valid a.s.

$$\overline{\lim}\,\frac{1}{N}\sum_{n=1}^{N} f(X_n) \le \int f_k d\mu + \|f\|_\infty \mu(G_k)$$

$$\underline{\lim}\,\frac{1}{N}\sum_{n=1}^{N} f(X_n) \ge \int f_k d\mu - \|f\|_\infty \mu(G_k).$$

Hence $\lim (1/N)\sum_{n=1}^{N} f(X_n) = \int f\, d\mu$ a.s. □

En lisant M. Lusin, on sera sans doute étonné d'apprendre que j'ai entre autre chose, inventé le procédé des cribles, construit le premier ensemble non analytique. Nul du moins ne sera plus étonné que je ne le fus. M. Lusin n'est parfaitement heureux que s'il a pu attribuer à quelque autre ses propres découvertes. Etrange manie; pardonnable il me semble, car il n'est guère à redouter que M. Lusin fasse école sur ce point.[4] [Henri Lebesgue, 1930]

C.6.4 Remarks

(1) Let $\mu = dx$ on $[0,1]^s$. The condition that (**) is satisfied for rectangles with rational vertices is equivalent to (X_n) being almost surely dx-distributed and is also equivalent to (*) being satisfied for all continuous f (Exercise).

(2) There exist almost surely μ-distributed sequences that are not Boreloid μ-distributed: Let ξ be a μ-distributed sequence and put $X_n = \xi_n$ for $n \ge 1$. Then (*) fails for

$$f = 1_{\{x:\exists n\ \xi_n = x\}}$$

(3) There exist Boreloid μ-distributed sequences: Let (X_n) be i.i.d. with common law μ.

Proposition C.6.5 Let $\xi = (\xi_n)$ be a uniformly distributed sequence on $[0,1]^s$, and let (V_n) be a sequence of i.i.d. random variables with absolutely continuous common law ν. Then the sequence $X_n = \{\xi_n + V_n\}$ is Boreloid uniformly distributed on the cube.

Proof We write $|A|$ for the Lebesgue measure of a Borel subset A of

[4]While reading M. Lusin, people will no doubt be astonished to learn that I have, among other things, invented the sieve method and constructed the first non-analytic set. Nobody, at least, will be more surprised than I was. M. Lusin is only completely happy when he can attribute his own discoveries to someone else. A strange habit; paradonable, it seems to me, since there is little chance that anyone will follow M. Lusin down that road.

the cube, and put

$$\mu_n(A) = \int_{\mathbb{R}^s} 1_A(\{\xi_n + x\}) d\nu(x).$$

Lemma C.6.6 For every Borel subset A,

$$\lim_{N\uparrow\infty} \frac{1}{N} \sum_{n=1}^{N} \mu_n(A) = |A|.$$

Proof of Lemma The function

$$g(y) = \int_{\mathbb{R}^s} 1_A(\{y + x\})\, d\nu(x)$$

is continuous, as it is the convolution product of a function in $L^\infty(\mathbb{R}^s)$ with a function in $L^1(\mathbb{R}^s)$. Hence

$$\frac{1}{N} \sum_{n=1}^{N} g(\xi_n) \to \int_{[0,1]^s} g(y)\, dy = |A|. \qquad \square$$

Now, to complete the proof of the proposition, we use a classical argument (which is used to prove a strong law of large numbers in the case of orthogonal variables in L^2). Let $\tilde{S}_N = (1/N) \sum_{n=1}^{N} (1_A(X_n) - \mu_n(A))$.

(a) $\tilde{S}_{N^2} \to 0$ a.s. Indeed,

$$\mathbb{P}(\exists N \geq m_1 : |\tilde{S}_{N^2}| > \epsilon) \leq \sum_{N=m_1}^{\infty} \frac{1}{\epsilon^2} \frac{1}{N^4} \sum_{n=1}^{N^2} [\mu_n(A) - (\mu_n(A))^2]$$

$$\leq \frac{1}{\epsilon^2} \sum_{N=m_1}^{\infty} \frac{1}{N^2},$$

and so $\overline{\lim}_N |\tilde{S}_{N^2}| \leq \epsilon$ a.s.

(b) $\tilde{S}_N \to 0$ a.s. Indeed, if $M^2 \le N < (M+1)^2$

$$|\tilde{S}_N - \tilde{S}_{M^2}| \le \left|\frac{N - M^2}{M^2}\right| + N\frac{2M}{NM^2} \le \frac{4}{M} \to 0.$$

So, by the lemma,

$$\frac{1}{N}\sum_{n=1}^{N} 1_A(X_n) \to |A| \qquad \text{a.s.},$$

and the result follows from Proposition C.6.2 . □

If the support of the measure ν is very narrow, the sequence (X_n) initially behaves like the sequence (ξ_n). Hence if (ξ_n) has a low discrepancy, the sequence (X_n) gives a fast integration as long as the required accuracy is not too high. For large n, however, the rate of integration deteriorates and the random character prevails.

C.6.7

Figure 2.3 shows what happens when (ξ_n) is taken to be a Halton sequence of dimension 1, 2, 3, 4, or 5 and when (V_n) is an i.i.d. sequence with uniform law on an interval of width 1, 0.1, 0.01, or 0.001.

We have computed an approximate discrepancy of the sequence $X_n = \xi_n + V_n$ and compared it with that of a Borel sequence (U_n) of the same dimension. The irregular graph is the logarithm to the base 2 of the ratio of these approximate discrepancies

$$\log_2 \frac{\text{App Disc}(X_n)}{\text{App Disc}(U_n)}.$$

When the width of the support of V_n is 1, the sequence (X_n) is i.i.d. By Exercise B.4.5. the graph is in a neighborhood of the strip

$$-\log\log\log n + \log\frac{\pi}{2} < . < \log\log\log n - \log\frac{\pi}{2}.$$

when the width of the support of V_n is progressively reduced from 0.1 to

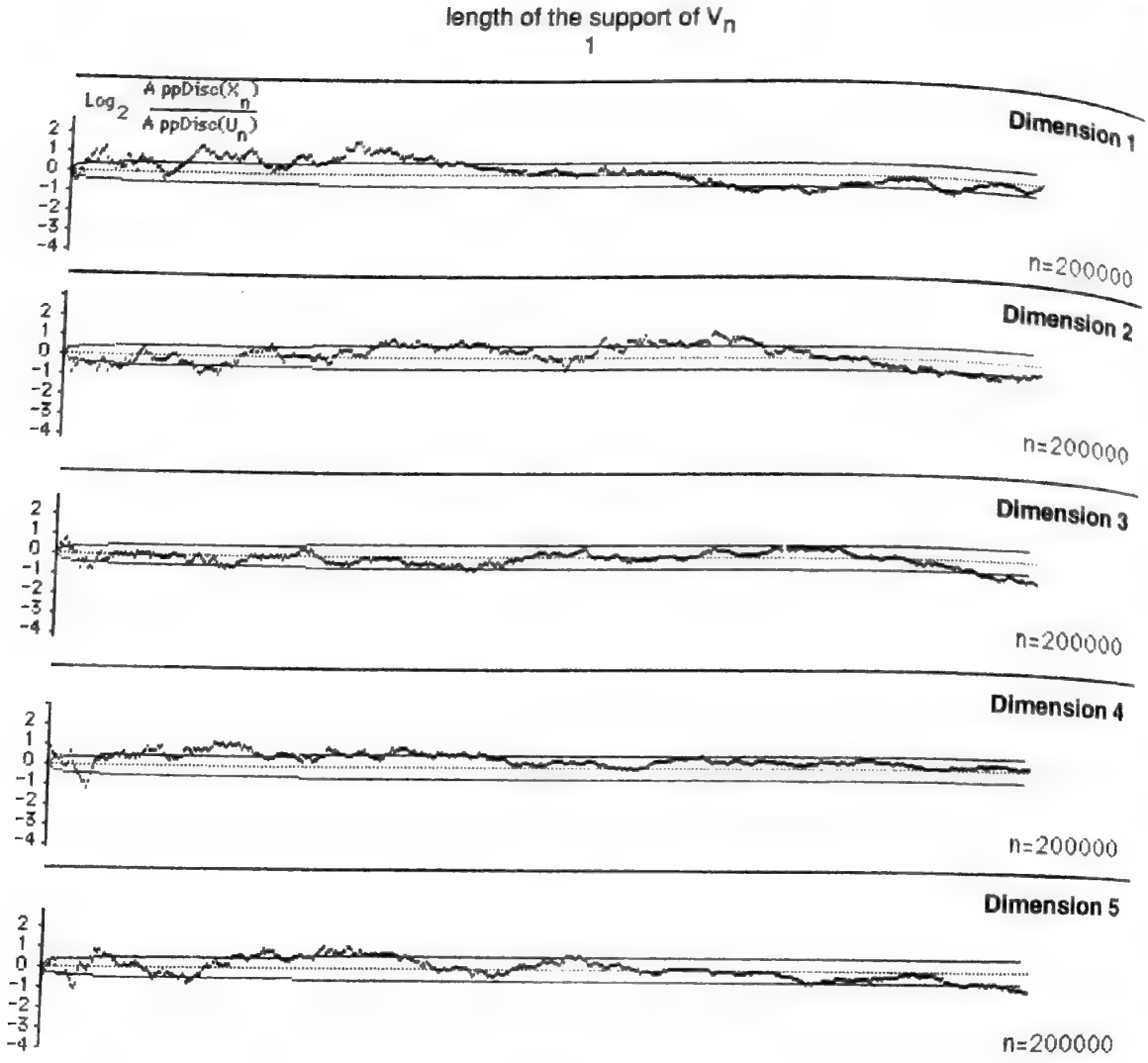


Figure 2.3

0.001, the approximate discrepancy of (X_n) is smaller than that of (U_n) (the logarithm of the ratio is negative). This becomes more significant as the support becomes smaller.

This phenomenon becomes weaker as the dimension increases, since for a fixed number of drawings (here 200,000) the quality of Halton sequences with respect to Borel sequences decreases when the dimension increases.

C.6.8 Remark

It is natural to ask whether it is possible in Proposition C.6.5 to take the V_n to be independent with law uniform on $[0, l_n]$ with $l_n \to 0$, in such a way that (X_n) remains Boreloid. We do not know the answer to this question. It is easily seen that the l_n must satisfy $\sum_n l_n = +\infty$.

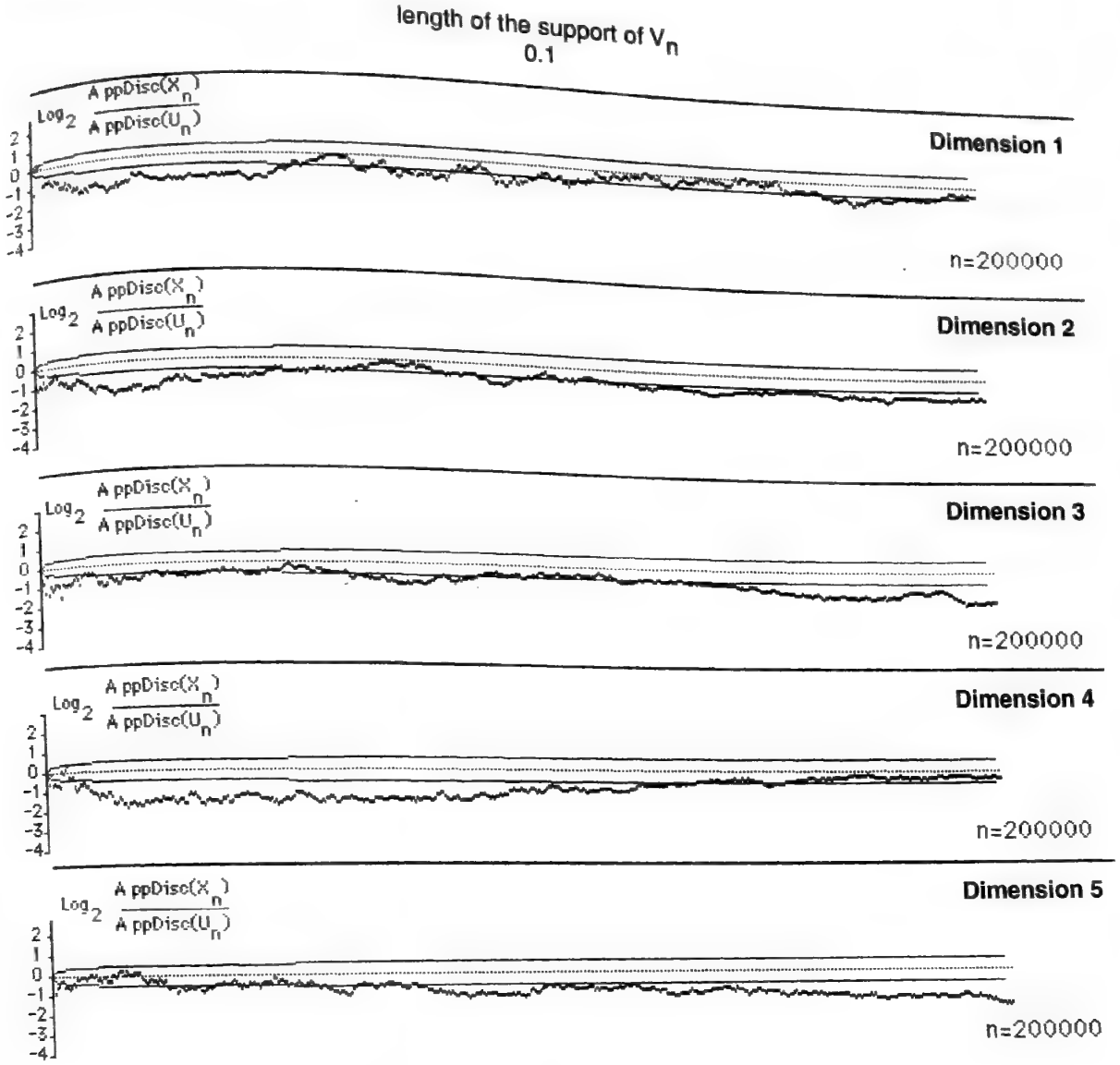


Fig. 2.3 (*Continued*)

D NUMERICAL COMPUTATION OF CONDITIONAL EXPECTATION

D.1 Complements on Conditional Expectation

D.1.1 Conditional Expectation Given a Random Variable

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. It is often useful to consider the σ-field $\sigma(Y)$ generated by a random variable Y: we will write $E(X|Y)$ in place of $E(X|\sigma(Y))$ for any integrable real random variable X, and call it the *conditional expectation of X given Y*. The following property is an easy consequence of the monotone class theorem.

Proposition D.1.2 Let Y be an $\mathbb{R}^d$-valued random variable and X an integrable real-valued random variable. There exists a Borel function $f: \mathbb{R}^d \to \mathbb{R}$ such that

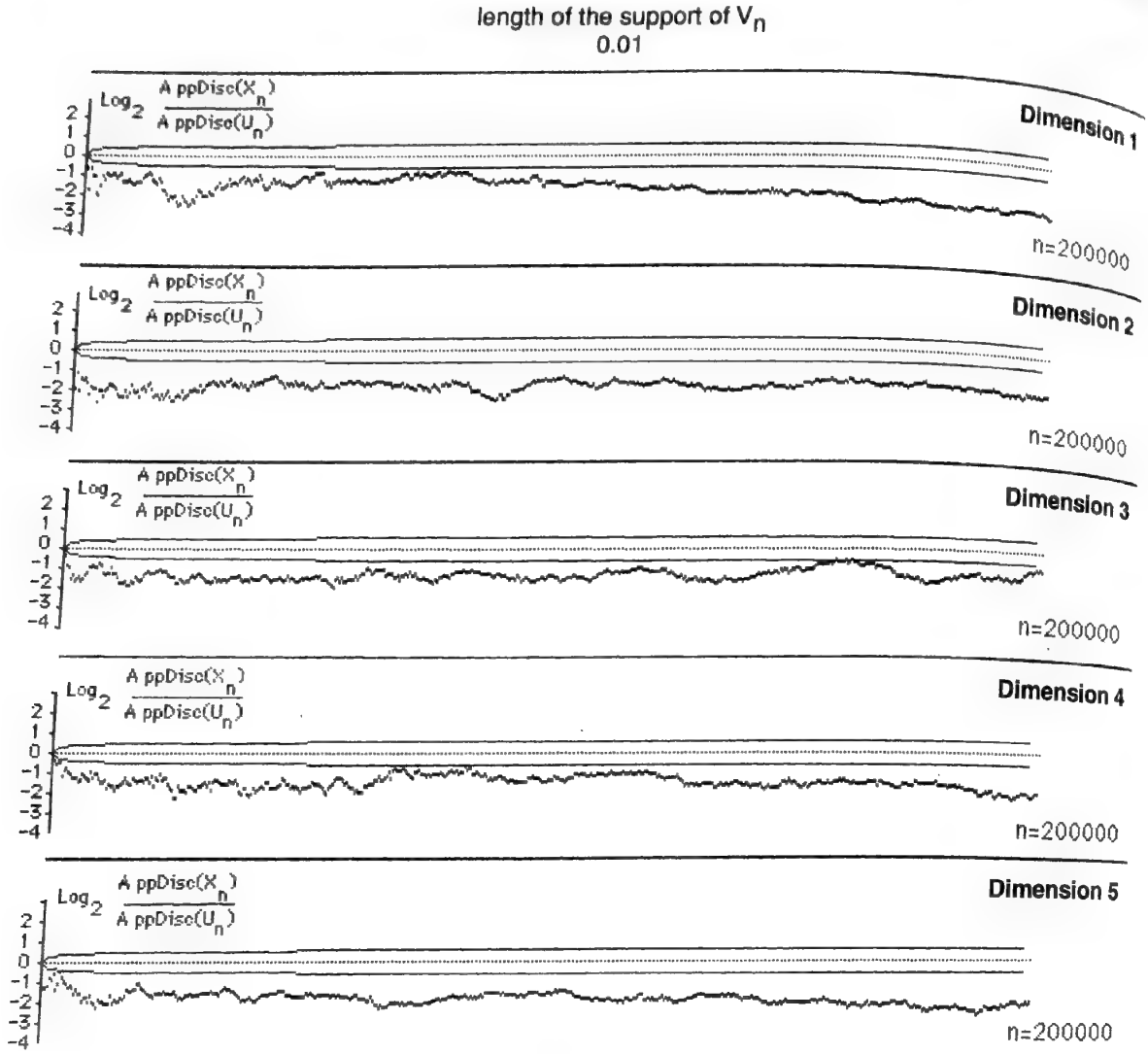


Figure 2.3 (*Continued*)

$$\mathbb{E}(X|Y) = f(Y) \qquad \mathbb{P} - \text{a.s.}$$

If we denote by $\mathbb{P}_Y$ the distribution of Y in $\mathbb{R}^d$, then the function f is unique up to a $\mathbb{P}_Y$-a.s. equality, and for any B in $\mathcal{B}(\mathbb{R}^d)$,

$$\int_{Y \in B} X \, d\mathbb{P} = \int_B f(y) \, d\mathbb{P}_Y(y).$$

The function f is often denoted by

$$f(y) = \mathbb{E}(X|Y = y),$$

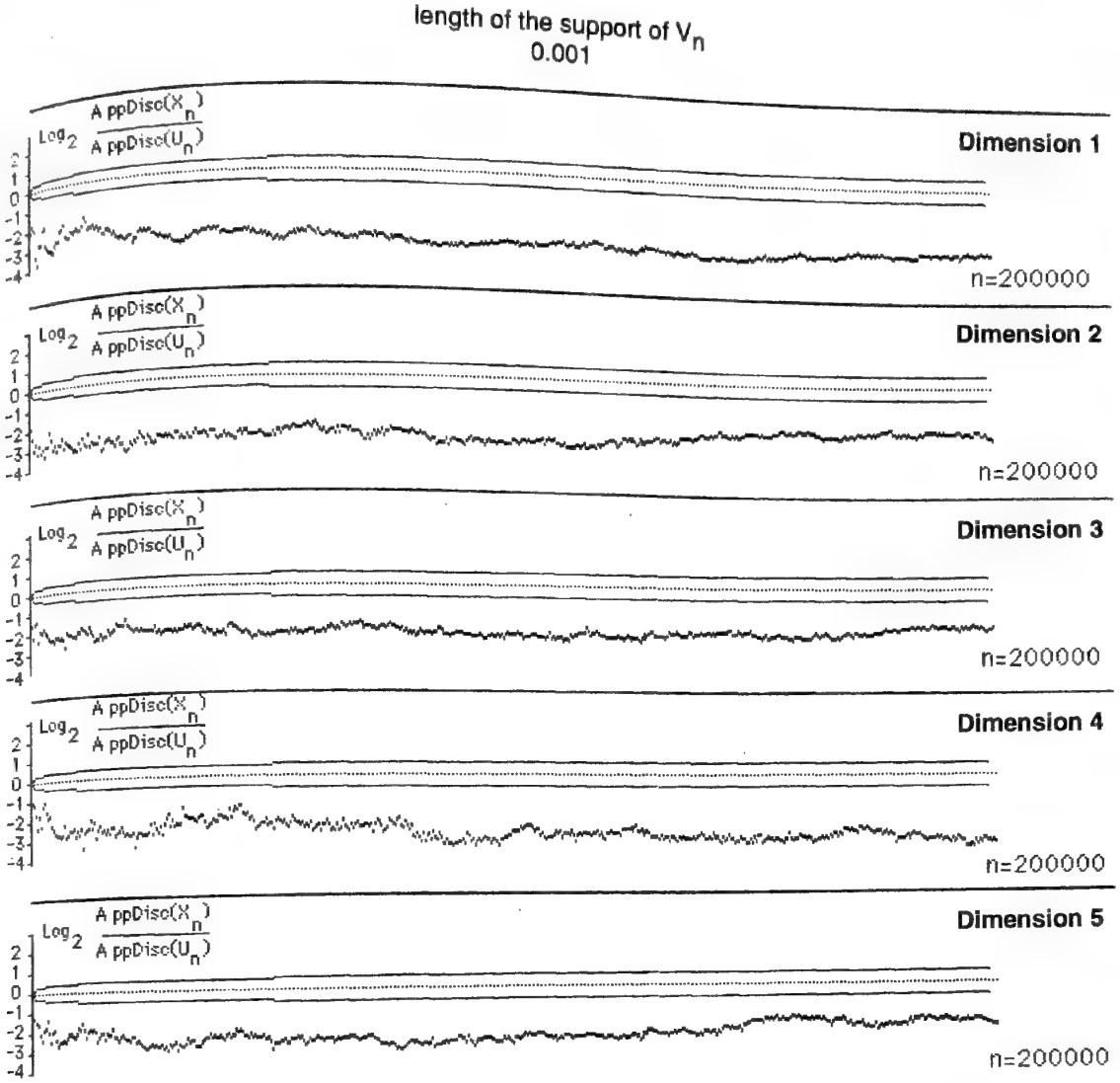


Figure 2.3 *(Continued)*

but this notation must be used carefully because the sets $\{Y = y\}$ may be $\mathbb{P}$-negligible.

Definition D.1.3 Let $(\Omega, .\mathcal{F})$ and $(E, \mathcal{E})$ be measurable spaces. A *Markov kernel* from $(\Omega, .\mathcal{F})$ into $(E, \mathcal{E})$ is a function $N: \Omega \times \mathcal{E} \to [0, 1]$ that satisfies the following properties:

(i) for fixed $\omega \in \Omega, C \rightsquigarrow N(\omega, C)$ is a probability on $(E, \mathcal{E})$;
(ii) for fixed $C \in \mathcal{E}, \omega \rightsquigarrow N(\omega, C)$ is $.\mathcal{F}$-measurable.

We have the following important existence theorem.

Theorem D.1.4 Let Y and X be as in Proposition D.1.2. There exists a Markov kernel N from $(\mathbb{R}^d, .\mathcal{B}(\mathbb{R}^d))$ into $(\mathbb{R}, .\mathcal{B}(\mathbb{R}))$ such that

(*) for any bounded Borel function φ on $\mathbb{R}$,

$$\mathbb{E}[\varphi(X)|Y] = \int \varphi(x)N(Y, dx) \qquad \mathbb{P}-\text{a.s.};$$

(**) for any bounded Borel function h on $\mathbb{R}\times\mathbb{R}^d$,

$$\mathbb{E}[h(X,Y)] = \int_{\mathbb{R}\times\mathbb{R}^d} h(x,y)N(y,dx)\mathbb{P}_Y(dy).$$

Such a kernel is called a *regular version of the conditional distribution of X given Y*. If (X,Y) has density ψ with respect to the Lebesgue measure on $\mathbb{R}\times\mathbb{R}^d$, a natural choice for N is

$$N(y,A) = \frac{\int_A \psi(x,y)\,dx}{\int_{\mathbb{R}} \psi(x,y)\,dx}$$

if $\int_{\mathbb{R}} \psi(x,y)\,dx > 0$ and $N(y,A) = 1_A(0)$ otherwise.

D.1.5 Numerical Computation of Conditional Expectation

Supppose we want to evaluate the function f in the proposition above. We are interested in the case when we know how to simulate the pair (X,Y) but either there is no explicitly computable density ψ or computing the ratio

$$f(y) = \frac{\int x\psi(x,y)\,dx}{\int \psi(x,y)\,dx}$$

will take a long time.

Methods are usually based on L^2-approximations, either by approximating f with explicitly computable smooth functions or by computing f point by point and storing these values.

D.2 Polynomial Approximation

First we assume that Y is a real-valued random variable.

Lemma D.2.1 If Y is a bounded random variable and if the support of its distribution is not a finite subset of the real line, then the random variables $1, Y, \ldots, Y^n, \ldots$, make up a linearly independent and total system in $L^2(\Omega, \sigma(Y), \mathbb{P})$.

Proof (a) From the assumption on the distribution of Y it follows that the equality

$$a_1 + a_1 Y + \cdots + a_n Y^n = 0 \qquad \text{a.s.}$$

leads to

$$a_0 = a_1 = \cdots = a_n = 0,$$

and this proves the linear independence.

(b) Put $K = \|Y\|_\infty$. It follows from the monotone class theorem that every random variable Z in $L^2(\Omega, \sigma(Y), \mathbb{P})$ may be written as $Z = f(Y)$ for some f in $L^2([-K, +K], \mathscr{B}([-K, +K]), \mathbb{P}_Y)$. Continuous functions are dense in that space, so for every $\epsilon > 0$ there exists some continuous function g such that

$$\|f - g\|_{L^2([-K,+K],\mathbb{P}_Y)} \leq \epsilon.$$

It follows from the Weierstrass theorem that there exists a polynomial P which meets

$$\sup_{y \in [-K,+K]} |g(y) - P(y)| \leq \epsilon,$$

and hence

$$\|Z - P(Y)\|_{L^2(\Omega,\sigma(Y),\mathbb{P})} = \|f - P\|_{L^2([-K,+K],\mathbb{P}_Y)} \leq 2\epsilon,$$

which completes the proof of the lemma. □

We shall only state the analogous result for an $\mathbb{R}^d$-valued random variable Y. First we need a definition.

Definition D.2.2 A $(d-1)$-dimensional *algebraic manifold* in $\mathbb{R}^d$ is the set of points $y = (y_1, \ldots, y_d)$ that satisfy the relation

$$P(y) = 0$$

for some polynomial P.

Lemma D.2.3 Suppose that Y is bounded and the support of its distribution is not included in any $(d-1)$-dimensional algebraic manifold. Then the system of random variables $\{Y_1^{n_1} \cdots Y_d^{n_d}, n_i \geq 0\}$ is linearly independent and total in $L^2(\Omega, \sigma(Y), \mathbb{P})$.

Suppose that the conditions of the lemma are satisfied. Let H_n be the C_{n+d}^d-dimensional vector space generated by the monomials $Y_1^{n_1} \ldots Y_d^{n_d}$, where $n_i \geq 0$ and $\sum_{i=1}^d n_i = n$, and let $P_n(Y)$ be the orthogonal projection of X onto H_n. As n tends to infinity, $P_n(Y)$ converges in $L^2(\Omega, \sigma(Y), \mathbb{P})$ to $\mathbb{E}(X|Y) = f(Y)$. To compute P_n, we must find the coefficients $a_{n_1,\ldots,n_d}$ that minimize

$$\left\| \sum_{n_1+\cdots+n_d \leq n} a_{n_1,\ldots,n_d} Y_1^{n_1} \cdots Y_d^{n_d} - X \right\|_{L^2}.$$

Let $\mathcal{Y}$ denote the column vector of the monomials $Y_1^{n_1} \cdots Y_d^{n_d}$ for some order in the index set $\{(n_1, \ldots, n_d) : n_1 + \cdots + n_d \leq n\}$. The column vector a of the coefficients $a_{n_1,\ldots,n_d}$ satisfies the equation

$$(*) \qquad \mathbb{E}[\mathcal{Y}\,\mathcal{Y}^t]a = \mathbb{E}[X\,\mathcal{Y}].$$

We must calculate the coefficients of the matrices $\mathbb{E}[\mathcal{Y}\,\mathcal{Y}^t]$ and $\mathbb{E}[X\,\mathcal{Y}]$, eventually by accelerated simulation. If we can simulate X and Y on a space whose dimension is not too large, then we can use low discrepancy sequences (Section C.5). The following is an attractive and direct method to solve equation (*).

D.2.4 Conjugate Gradient Algorithm

We want to solve the system

$$(*) \qquad Ax = b$$

where b is in $\mathbb{R}^p$ and A is a $p \times p$ positive-definite symmetric matrix. The problem can be considered as a minimization problem for the function

$$J(y) = \tfrac{1}{2}\langle Ay, y\rangle - \langle b, y\rangle.$$

The conjugate gradient method works by computing a finite sequence of vectors $x_0, x_1, \ldots, x_p$ and directions $w_0, w_1, \ldots, w_p$ such that w_n is the search direction of x_{n+1} from x_n and J is minimal at the point x_{n+1} along this line; the new search direction w_{n+1} will be the conjugate direction of w_n at the point x_{n+1}. More explicitly, by setting

$$g_n = Ax_n - b,$$

the algorithm goes as follows:

(1) the starting vector x_0 is arbitrary, and $w_0 = g_0$

(2) $x_{n+1} = x_n + \rho_n w_n$, where the scalar ρ_n is chosen so that

$$J(x_n + \rho_n w_n) \leq J(x_n + \theta w_n) \qquad \text{for any } \theta \text{ in } \mathbb{R},$$

(3) $w_{n+1} = g_{n+1} + \alpha_{n+1} w_n$, where the scalar α_{n+1} is chosen so that

$$(w_{n+1}, Aw_n) = 0$$

As we shall see, if $w_n \neq 0$, then

(4)

$$\rho_n = -\frac{(g_n, w_n)}{(Aw_n, w_n)}$$

and

(5)

$$\alpha_{n+1} = -\frac{(g_{n+1}, Aw_n)}{(Aw_n, w_n)}.$$

We have the following result.

Theorem D.2.5 The algorithm defined by (1), (2), (3), (4), and (5) converges in at most p iterations to the solution x of $Ax = b$.

Proof (a) Notice first that if $g_n = 0$ then x_n is the solution.
(b) In step (2), we have

$$\frac{\partial}{\partial \rho} J(x_n + \rho w_n) = 0$$

for $\rho = \rho_n$, and hence (4) holds.

(c) In step (3),

$$(g_{n+1} + \alpha_{n+1} w_n, A w_n) = 0,$$

and hence (5) holds.

(d) We will show by induction that, for any $n \geq p$ such that $g_n \neq 0$, we have for any $i < n$

$$\begin{aligned} (g_n, g_i) &= 0 \\ (g_n, w_i) &= 0 \\ (w_n, A w_i) &= 0 \\ \alpha_n &\neq 0 \\ w_n &\neq 0 \\ \rho_n &\neq 0 \end{aligned}$$

(e) If $n = 0$, the first four statements are obvious; we have $w_0 = g_0 \neq 0$ (otherwise x_0 is the solution) and

$$\rho_0 = \frac{(g_0, w_0)}{(A w_0, w_0)} \neq 0.$$

(f) Suppose that the statements hold for n. If $g_{n+1} \neq 0$, we have

- $$\begin{aligned} (g_{n+1}, w_n) &= (A x_{n+1} - b, w_n) \\ &= (A(x_n + \rho_n w_n) - b, w_n) \\ &= (g_n, w_n) + \rho_n (A w_n, w_n) \\ &= 0 \qquad \text{by (4)} \end{aligned}$$

- for $i < n$,

$$\begin{aligned}(g_{n+1}, w_i) &= (g_{n+1} - g_n, w_i) && \text{by the inductive hypothesis}\\ &= \rho_n(Aw_n, w_i)\\ &= 0 && \text{by the inductive hypothesis}\end{aligned}$$

- for $i < n+1$,

$$\begin{aligned}(g_{n+1}, g_i) &= (g_{n+1}, w_i - \alpha_i w_{i-1})\\ &= 0\end{aligned}$$

- for $i < n$,

$$\begin{aligned}(w_{n+1}, Aw_i) &= (g_{n+1}, Aw_i) + \alpha_{n+1}(w_n, Aw_i)\\ &= \left(g_{n+1}, \frac{g_{i+1} - g_i}{\rho_i}\right)\\ &= 0\end{aligned}$$

-

$$(w_{n+1}, Aw_n) = 0 \qquad \text{by the choice of } \alpha_{n+1}$$

-

$$\begin{aligned}\alpha_{n+1} &= -\frac{(g_{n+1}, Aw_n)}{(Aw_n, w_n)}\\ &= -\frac{(g_{n+1}, g_{n+1} - g_n)}{(g_{n+1} - g_n, w_n)}\\ &= -\frac{|g_{n+1}|^2}{(g_{n+1} - g_n, g_n + \alpha_n g_{n-1})}\\ &= \frac{|g_{n+1}|^2}{|g_n|^2}\\ &> 0\end{aligned}$$

-

$$\begin{aligned}|w_{n+1}|^2 &= (g_{n+1} + \alpha_{n+1}g_n, g_{n+1} + \alpha_{n+1}g_n)\\ &= |g_{n+1}|^2 + (\alpha_{n+1})^2|g_n|^2\\ &> 0\end{aligned}$$

$$\begin{aligned}\rho_{n+1} &= -\frac{(g_{n+1}, w_{n+1})}{(Aw_{n+1}, w_{n+1})} \\ &= -\frac{|g_{n+1}|^2}{(Aw_{n+1}, w_{n+1})} \\ &< 0.\end{aligned}$$

(g) The $p+1$ directions g_n, for $0 \le n \le p$, are orthogonal. Therefore $g_n = 0$ for some $n \le p$. This completes the proof of the theorem. □

D.2.6 Remarks

(1) If p is large,we stop the process as soon as the ratio $|g_{n+1}|^2/|g_0|^2$ is small enough.

(2) We are well advised to compute α_n with double precision.

D.2.7 Particular Cases

We can often use classical orthogonal polynomials, which avoids computation.

For instance, consider the case of Legendre polynomials. The Legendre polynomials are defined by formulas

$$Le_0(y) = 1$$

$$Le_n(y) = (-1)^n \frac{\sqrt{2n+1}}{n!} \frac{d^n}{dy^n}[(1-y)^n y^n], \qquad (n = 1, 2, \ldots).$$

The collection

$$Le_{\vec{n}}(x) = Le_{n_1}(x_1) \cdots Le_{n_d}(x_d), \qquad x = (x_1, \ldots, x_d), \quad \vec{n} = (n_1, \ldots, n_d)$$

is an orthonormal basis in $L^2([0,1]^d, \lambda_d)$. Suppose that X and Y can be simulated by

$$\begin{aligned}X &= f_1(U, W) \\ Y &= f_2(V, W)\end{aligned}$$

and (U, V, W) is a u.d. random vector in $[0,1]^{s_1} \times [0,1]^{s_2} \times [0,1]^{s_3}$. Suppose also that the conditional distribution N of V given Y is easily calculated. Here the desired function $f(y) = \mathbb{E}(X|Y = y)$ takes the value

$$f(y) = \int N(y, dv)h(v)$$

where $h(v) = \mathbb{E}[X|V = v]$ (exercise: prove it). So we must compute the function h, which can be done by computing (using simulation or accelerated simulation) the scalar products

$$a_{\vec{n}} = \mathbb{E}[XLe_{\vec{n}}(V)].$$

In $L^2([0,1]^{s_2}, \lambda_{s_2})$, we use the expansion

$$h(v) = \lim_{k \uparrow \infty} \sum_{|\vec{n}| \le k} a_{\vec{n}} Le_{\vec{n}}(v)$$

and the final step is to compute the integrals

$$\int N(y, dv)Le_{\vec{n}}(v).$$

The Chebyshev polynomials

$$T_0(y) = 1$$
$$T_n(y) = \sqrt{2}\cos n\theta \qquad \text{for} \quad y = \cos\theta$$

are an orthonormal basis in $L^2([-1,+1], dy/(\pi\sqrt{1-y^2}))$. They are especially useful for problems involving uniformly distributed random variables on the unit circle.

The Laguerre polynomials

$$L_n(y) = \frac{e^y}{n!}\frac{d^n}{dy^n}(y^n e^{-y}), \qquad n \ge 0$$

are an orthonormal basis in $L^2([0,\infty[, e^{-y}dy)$. They are convenient when exponentially distributed random variables are involved.

The Hermite polynomials

$$\Psi_n(y) = \frac{(-1)^n}{\sqrt{n!}} e^{y^2/2}\frac{d^n}{dy^n}[e^{y^2/2}], \qquad n \ge 0$$

are an orthonormal basis in $L^2(\mathbb{R}, N(0,1))$. They are useful when Y is a Gaussian variable, but X or the pair (X, Y) are not.

D.3 Point by Point Calculation

In order to deal with the rejection method, we assume that the pair (X, Y) can be simulated on a Polish probability space $(W, \mathcal{W}, \mu)$ in which we know a sequence $\xi = (\xi_n)$ such that

$$\frac{1}{N}\sum_{n=1}^{N} \delta_{\xi_n}$$

converges narrowly to μ. For example, W may be $[0,1]^d$ or $[0,1]^{\mathbb{N}}$. For the sake of simplicity we assume that Y is a $[0,1]$-valued random variable (it is easy to extend the results to $[0,1]^s$ or $\mathbb{R}^s$).

In practice the real number $Y(\omega)$ is represented by some binary digits and we compute the function $f(y) = \mathbb{E}[X|Y = y]$ at the points of this dyadic network. Let

$$x = \sum_{i=1}^{\infty} \frac{e_i}{2^i}$$

be the canonical dyadic expansion of $x \in [0,1[$; that is, $e_i = 0$ or 1 and infinitely many of the e_i are zero. For $k \geq 1$, consider the truncation map T_k given by

$$x \overset{T_k}{\longmapsto} x_k = \sum_{i=1}^{k} \frac{e_i}{2^i} = \frac{[2^k x]}{2^k}.$$

We can now state the following approximation result.

Lemma D.3.1 Let X be an integrable real random variable, and put $Y_k = T_k(Y)$. Then,

$$\mathbb{E}[X|Y] = \lim_{k \uparrow \infty} \mathbb{E}[X|Y_k]$$

a.s. and in L^1-norm. For $p \geq 1$, if X belongs to L^p, then convergence holds also in L^p-norm.

Proof Let $\mathscr{P}_k$ be the partition $\{[i/2^k, (i+1)/2^k[, i = 0, \ldots, 2^k - 1\}$ and let $\tilde{\mathscr{P}}_k$ be the related algebra. It follows from the equality

$$\bigvee_k \tilde{\mathscr{P}}_k = \mathscr{B}([0, 1[)$$

that

$$\bigvee_k \sigma(Y_k) = \sigma(Y).$$

The proof is completed by applying the martingale convergence theorem. □

Let us restate the conclusion in different terms: if

$$\mathbb{E}[X|Y_k] = f_k(Y_k),$$

then

$$f(Y) = \lim_k f_k(Y_k) = \lim_k f_k \circ T_k(Y) \qquad \text{a.s.}$$

So we must compute the functions f_k. We use the following method.

(i) Take a point ξ_n from the sequence ξ, and compute $Y(\xi_n)$ and $X(\xi_n)$.
(ii) For $k \in \mathbf{N}$, and $i = 0, \ldots, 2^k - 1$, compute the ratio m_i given by the sum of the values of $X(\xi_n)$ such that $Y(\xi_n) \in [i2^{-k}, (i+1)2^{-k}[$ divided by the number of such $Y(\xi_n)$.
(iii) Set

$$f_k(i2^{-k}) = m_i.$$

In fact, it is clear that

$$\mathbb{E}[X|Y_k] = f_k(Y_k) = \sum_{i=0}^{2^k-1} \frac{\mathbb{E}[X 1_{[Y \in [i2^{-k}, (i+1)2^{-k}[]}]}{\mathbb{P}(Y \in [i2^{-k}, (i+1)2^{-k}[)} 1_{[i2^{-k}, (i+1)2^{-k}[}(Y_k),$$

and we have

$$\frac{\mathbb{E}[X1_{[i2^{-k},(i+1)2^{-k}[}(Y)]}{\mathbb{P}(Y \in [i2^{-k},(i+1)2^{-k}[)} = \lim_{N\uparrow\infty} \frac{\sum_{n=1}^{N} X(\xi_n)1_{[i2^{-k},(i+1)2^{-k}[}(Y(\xi_n))}{\sum_{n=1}^{N} 1_{[i2^{-k},(i+1)2^{-k}[}(Y(\xi_n))}$$

whenever the functions X and $1_{[i2^{-k},(i+1)2^{-k}[} \circ Y$ are μ-Riemann integrable. In particular, this last condition holds under the hypothesis that X and Y are μ-Riemann integrable and the set of dyadic rationals is negligible for the probability law of Y.

The following algorithm works when the pair (X, Y) is simulated in $[0, 1]^d$, and the sequence (ξ_n) is given by a regular network of 2^{md} points from $[0, 1]^d$. The function f_k is stored in a file with 2^k entries and the accuracy to which f is defined is controlled by the number k. (See Figure 2.4.)

```
Program EspXgivenY;                          {computation of E[X|Y = y] = f(y) when the pair
                                              (X, Y) is simulated on [0, 1]^d}

const  d=10; k=10; m=10;
       kk=1024; mm=1024;                     {kk = 2^k and mm = 2^m}
var    a: array [0..kk] of real;
       b: array [0,,kk] of integer;          {the vector a will contain the values of f and the
                                              vector b will contain the number of values which
                                              are averaged so that the precision can be given}

procedure ConditionalExp;
type vector = array [1..d] of real;
var    v:vector;
       r: integer;

       procedure step {u:vector};
       var i:integer;x,y:real;
       begin
          % computation of
             y:=Y(u[1], ..., u[d]) %
          % computation of
             x:=X(u[1], ..., u[d]) %
          i:=trunc(kk*y);                    {what gives the index i such that y be in the
                                              interval [i/2^k, (i+1)/2^k[}

          b[i]:=b[i]+1;
          a[i]:= a[i]+(x-a[i])/b[i];         {updating of the average of the X(u)'s for the
                                              u's such that Y(u) be in [i/2^k, (i+1)/2^k[}

       procedure unwind(q:integer);          {recursive procedure}
       var j:integer;
```

Figure 2.4

D.3.2 Remarks

(1) The function f is not calculated at points y that are not visited by Y and it is inaccurate at points that are seldom visited; it is then interesting to store, in addition to the file which gives f_k, a file giving the number of computed values $c[i]$. This provides some information about the accuracy obtained.

(2) Would it be more profitable to take for (ξ_n) a low discrepancy sequence in $[0,1]^d$? It depends very much on the desired accuracy.

(3) It is important to note an essential difference from polynomial methods, which often work much faster: the method above keeps up positivity. The conditional expectation of a positive random variable is positive, and so is the numerical approximation obtained by the point-by-point method. This may be imperative in some applications, for instance when X is a physical quantity like dissipated energy, mass, or inertia.

```
        begin                                  {let u run through the discretization of [0,1]^d
            if q=d+1 then step(v)              and does step(u) of each time}
            else for j:=1 to mm do
            begin
                v[q]:=j/mm;
                unwind(q+1)
            end
        end;
begin
        for r:=0 to kk do
        begin
            a[r]:=0; b[r]:=0;
        end;
        unwind(1)
end;

function f(y:real): real;
var i: integer;
begin
        i:=trunc(kk*y);
        f:=a[i]
end;

begin                                          {main program}
ConditionalExp;
% use of the function f%
end.
```

Figure 2.4 (*Continued*)

D.4 Comparison of Both Methods on an Example

D.4.1 Principle of Dynamical Weighing

Suppose we want to take measurements from vehicles without stopping them in order to assess whether or not the rules about maximum static loads per axle-tree are being kept. A device (piezoelectric cable) is set on the road and, following calibration, evaluates with suitable accuracy the real force undergone by the ground when an axle-tree runs over. The movement of the vehicle (which depend on road irregularities, braking, acceleration, and so on) may lead to significant differences between the load at rest per axle-tree and the measured quantity.

Therefore we want to estimate as closely as possible the load at rest per axle-tree from the measurements recorded by one or several piezoelectric bars.

D.4.2 Conditional Expectations to Be Computed

A vehicle can be represented by

- the vector of static loads per axle-tree

$$\vec{M} = (m_1, m_2, \ldots, m_r), \qquad r = \text{the number of axle-trees}$$

- the vector of parameters

$$\vec{\Lambda} = (\lambda_1, \ldots, \lambda_n),$$

which specify the geometrical and mechanical features.

In the case when three bars are embedded in the ground, the weight measurements are displayed by three vectors

$$\vec{P}_1 = (p_{11}, \ldots, p_{1r})$$
$$\vec{P}_2 = (p_{21}, \ldots, p_{2r})$$
$$\vec{P}_3 = (p_{31}, \ldots, p_{3r})$$

Some features are easily calculated, such as the velocity (since there is more than one bar) and therefore the distance between axle-trees. Other quantities cannot be explicitly deduced from the measurements, so we must choose a reasonable distribution for them. Thus we give a probability distribution to the pair $(\vec{M}, \vec{\Lambda})$ and hence endow the vectors $(\vec{M}, \vec{P}_1, \vec{P}_2, \vec{P}_3)$ with a joint distribution. The problem is then to compute the conditional

expectation,

$$\mathbb{E}[\vec{M}|\vec{P}_1 = \vec{p}_1, \vec{P}_2 = \vec{p}_2, \vec{P}_3 = \vec{p}_3] = f(\vec{p}_1, \vec{p}_2, \vec{p}_3),$$

and possibly the conditional variance

$$\mathbb{E}[|\vec{M} - f(\vec{p}_1, \vec{p}_2, \vec{p}_3)|^2 | \vec{P}_1 = \vec{p}_1, \vec{P}_2 = \vec{p}_2, \vec{P}_3 = \vec{p}_3].$$

> Models are, for the most part, caricatures of reality, but if they are good, then, like good caricatures, they portray, though perhaps in distorted manner, some features of the real world. [Mark Kac, 1969]

D.4.3 Application of the Polynomial Method

We write X and Y in place of $\vec{M}$ and $(\vec{P}_1, \vec{P}_2, \vec{P}_3)$. We may assume that X is real-valued, since the conditional expectation is computed component by component. We make the assumptions of lemma D.2.3.

(a) *Computer representation of the monomials.* The C^n_{n+d} components of $\mathcal{Y}$ must be organized in a linear order by some convenient procedure. For this purpose the vector of C^n_{n+d} multiindices is built up by a twofold recurrence in n and d. This recurrence takes the form of an array of pointers, which is built and run through very quickly.

(b) *Matrix and covariance vector calculation.* The matrix $\mathbb{E}[\mathcal{Y}\mathcal{Y}^t]$ and the vector $\mathbb{E}[X\mathcal{Y}]$ are evaluated by the classical Monte Carlo method: after each drawing of Y and X a new copy of the matrix $\mathcal{Y}\mathcal{Y}^t$ and the vector $X\mathcal{Y}$ is computed and the mean is updated.

(c) *Implementation of the conjugate gradient-algorithm.* The matrix $\mathbb{E}[\mathcal{Y}\mathcal{Y}^t]$ is positive semi-definite and so are its Monte Carlo approximations. However, the approximation may not be a positive definite matrix. The conjugate gradient algorithm provides a minimum value for the quadratic form, but this minimum is not unique and depends on the arbitrary initial value. In practice, this (nonunique) value is nevertheless acceptable as an approximation of the conditional expectation.

(d) *Output.* Once we have run the program, we can compute the conditional expectation of the weight per axle-tree given the measurements. This is achieved quickly and accurately because the conditional expectation has a polynomial expression with respect to the recorded data.

D.4.4 Comparison with the Point-by-Point Method

The two methods give the same results, but the point-by-point method, while easier to use, has some intrinsic drawbacks:

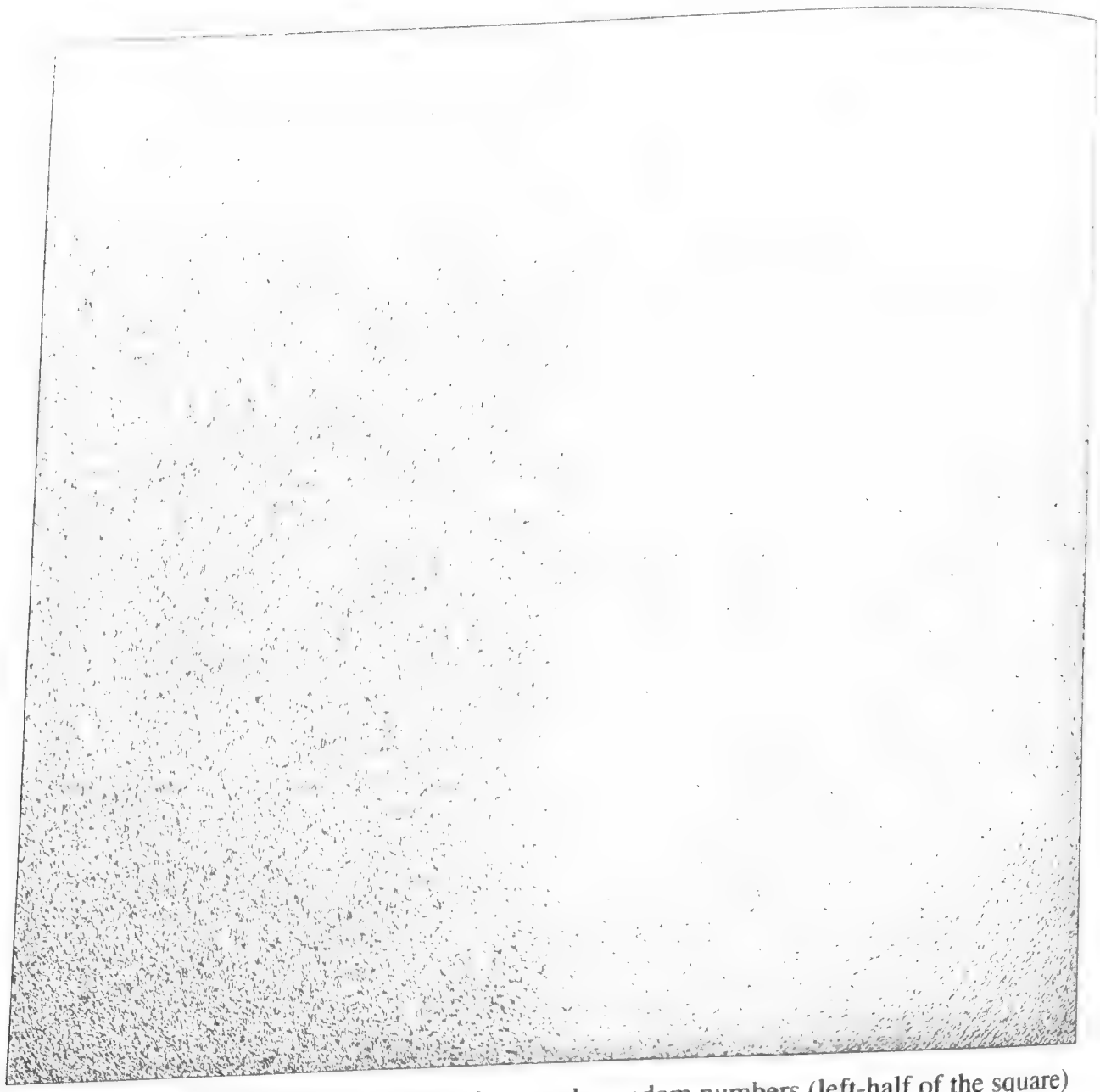


Figure 2.5 Gradation of grey obtained by pseudo-random numbers (left-half of the square) and by a low-discrepancy sequence (right-half of the square). Although the same number of points are thrown into the two halves, the grey appears darker on the right-hand side because the small black spots overlap more often on the left-hand side.

- the way we choose our discrete set has a strong effect on the result;
- if the random variable Y seldom takes values near y, then the estimation of $\mathbb{E}[X|Y = y]$ is very bad, and outliers are observed.

Figure 2.5 is an illustration of the discussion in Section C.5: in low dimensions, low-discrepancy sequences are more regularly distributed than pseudo random sequences.

CHAPTER 3

Simulation of Random Processes

In this chapter we deal with simulation methods for computing quantities related to the main classes of random processes: stationary, Markov, and point processes. Infinite-dimensional representations are quite usual here and therefore we must proceed with care and make sure that our definitions are rigorous.

A INTEGRATION IN LARGE OR INFINITE DIMENSIONS

In this section we formalize the main notions we shall need in infinite dimensions and introduce the Bernoulli shift method.

A.1 Changing Dimension

Let $(\Omega_1, \mathscr{A}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathscr{A}_2, \mathbb{P}_2)$ be Polish spaces endowed with their Borel σ-fields and probability measures $\mathbb{P}_1$, $\mathbb{P}_2$. Let φ be a measurable map from Ω_1 into Ω_2 such that

(i) $\mathbb{P}_1\{x \in \Omega_1 : \varphi \text{ is not continuous at } x\} = 0$;
(ii) $\varphi_* \mathbb{P}_1 = \mathbb{P}_2$.

If f is a $\mathbb{P}_2$-Riemann integrable function, then $f \circ \varphi$ is $\mathbb{P}_1$-Riemann integrable and if $\xi = (\xi_n)$ is $\mathbb{P}_1$-distributed, we have

$$\frac{1}{N} \sum_{n=1}^{N} f \circ \varphi(\xi_n) \to \int f \circ \varphi \, d\mathbb{P}_1 = \int f \, d\mathbb{P}_2.$$

It follows that $\varphi(\xi_n)$ is $\mathbb{P}_2$-distributed.

A.1.1 Example

For $x \in]0, 1]$ let

$$x = \sum_{n=1}^{\infty} \frac{a_n}{2^n}, \qquad a_n \in \{0, 1\}$$

be the dyadic expansion of x such that $\sum_{n=1}^{\infty} a_n = +\infty$. Define the mapping $\varphi_s :]0, 1] \to [0, 1]^s$ by

$$\varphi_s(x) = (y_1, \ldots, y_s),$$

where

$$y_p = \sum_{n=1}^{\infty} \frac{a_{sn+p-1}}{2^n} \qquad p = 1, \ldots, s.$$

Then φ_s satisfies (i) and (ii) with $\mathbb{P}_1 = \lambda_1$ and $\mathbb{P}_2 = \lambda_s$.

A.1.2 Remark

It would be interesting to use mappings φ that transform low-discrepancy sequences in $[0, 1]$ into low discrepancy sequences in $[0, 1]^s$ or $[0, 1]^{\mathbb{N}}$. However, such good mappings are not known, and the dimension of the canonical space has an important role in simulation. It does matter whether a random variable is defined on $[0, 1]$ or on $[0, 1]^{20}$.

A.2 Simulation of a Non-Riemann Integrable Random Variable

Let X be a non-Riemann integrable random variable defined on the space $([0, 1], \mathscr{B}([0, 1]), \lambda_1)$. It is not possible to use the variable X directly in order to approximate the law of X by simulation. If (ξ_n) is u.d. then the average $N^{-1} \sum_{n=1}^{N} X(\xi_n)$ does not in general converge to $\mathbb{E}(X)$.

A.2.1 Example

Let $x \in]0, 1]$ be given by

$$x = \sum_{n=1}^{\infty} \frac{a_n}{2^n}, \qquad a_n \in \{0, 1\} \quad \text{with} \quad \sum_{n=1}^{\infty} a_n = +\infty.$$

Put

$$X(x) = \sup_{n\geq 1} \frac{1}{n} \sum_{k=1}^{n} a_k.$$

Then X is not Riemann integrable. Indeed, let

$$A = \left\{ x \in]0, 1] : \lim_n \frac{1}{n} \sum_{k=1}^{n} a_k = \frac{1}{2} \text{ and } a_1 = 0 \right\}.$$

We have $\lambda_1(A) = \frac{1}{2}$ and $X(x) < 1$ for any $x \in A$, whereas in every neighborhood of x there exists y that satisfies $X(y) = 1$.

A.2.2 Example

Let (X_n) be the sequence of coordinate mappings of $([-1, +1], \mathscr{B}([-1, +1]), dx/2)^{\otimes \mathbf{N}^*}$. Put

$$S_n = X_1 + \cdots + X_n$$

$$X = \left(\left(\sup_{n\geq 3} \frac{S_n}{(2n \log \log n)^{1/2}} \right) \wedge 1 \right) \vee 0.$$

It can be proved that X is a bounded non-Riemann integrable variable.

Definition A.2.3 A non Riemann integrable random variable X is said to be *simulated* if it is possible for some $s \in \mathbf{N}^* \cup \{+\infty\}$ to exhibit a Riemann integrable random variable Y defined on $([0, 1]^s, \mathscr{B}([0, 1]^s), \lambda_s)$ with the same law as X.

It is easier to find such a variable than to obtain an analytic expression of the distribution of X (here analytic simply means: involving only standard functions).

A.2.4 Example: Polya Urn Scheme

Suppose that we have an urn that initially contains w white and b black balls. At each draw we remove a ball at random and then replace it, together with another ball of the same color. Let W_n and B_n denote the number of white and black balls, respectively, in the urn after the nth drawing, and write $x_n = W_n/(W_n + B_n)$, $n \geq 0$, for the proportion of white balls. Note also that $W_n + B_n = w + b + n$. A representation of such an experiment is given by introducing a Borel sequence $(U_n)_{n\geq 1}$ on $[0, 1]$ and writing

$$x_0 = w/(w+b)$$

$$x_{n+1} = \frac{1}{w+b+n+1}\left((w+b+n)x_n + 1_{\{U_{n+1}<x_n\}}\right) \qquad n \geq 0.$$

The sequence (x_n) turns out to be a Markov chain. It is also clear that (x_n) is a bounded martingale with respect to the past history σ-fields $\mathcal{F}_n = \sigma(U_k; 1 \leq k \leq n) = \sigma(x_k; 1 \leq k \leq n)$, and that the martingale convergence theorem therefore is applicable. The limit random variable

$$x_\infty = \lim_n \sup x_n,$$

a mapping from $[0, 1]^{\mathbf{N}}$ equipped with the product topology into $[0, 1]$, is not Riemann integrable with respect to the product measure $dx^{\otimes \mathbf{N}}$, but if $w = b = 1$, its distribution is uniform in $[0, 1]$, and so x_∞ is simulated.

A.2.5 Exercise

(1) Show that in the preceding example, the law of x_∞ is in general a beta-law with density

$$(1-x)^{w-1}x^{b-1}\frac{(w+b-1)!}{(w-1)!(b-1)!}.$$

(2) Show that if the urn has initially three balls of different colors, the barycentric representation of the ratio of balls of each color converges almost surely to a random point in the equilateral triangle with unit sides, which is distributed in proportion to the area measure.

A.3 Effective Computation of the Expectation of Random Variables in L^1

Unbounded random variables are very commonly used in simulation. For instance, the log function on $]0, 1]$ is used to simulate the exponential law by

$$X = -\frac{1}{\lambda}\log U.$$

But unbounded random variables cannot be used if the position of poles is unknown. Even when the position is known, the result will depend on the speed with which the uniformly distributed sequence nears the poles.

A.3.1 Example

To compute $\int_0^1 x^{-1/2}\,dx = 2$ by simulation, let us use a u.d. sequence $\xi = (\xi_n)$ in $[0, 1]$. If the sequence ξ is changed on a null set $A \subset \mathbf{N}^*$ (i.e., such that $N^{-1}\sum_{n=1}^{N} 1_A(n) \to 0$ as N goes to infinity), the new sequence is still u.d. Let $\xi_{p^2} = p^{-8}$ for $p \geq 1$; then

$$\limsup_n \frac{1}{N}\sum_{n=1}^{N} \frac{1}{\sqrt{\xi_n}} = +\infty. \qquad \square$$

Nevertheless, it is often possible to compute the expectation of unbounded or even non-Riemann integrable functions without proceeding to a reduction to the Riemann integrable case.

A.3.2 Example

Let $(U_n)_{n\geq 1}$ be the Borel sequence obtained by considering the coordinate mappings of $([0, 1]^\infty, \lambda_\infty)$. Put

$$X_n = \sum_{i=1}^{n} \frac{2U_i - 1}{i} \qquad \text{and} \qquad Y = \limsup_n |X_n|.$$

It can be proved that Y is not bounded and that $(Y \wedge 1) \vee (-1)$ is not Riemann integrable. The martingale X_n, however, converges a.s. and in quadratic mean, and we have the estimate

$$\|Y - |X_n|\|_{L^2}^2 \leq \sum_{i=n+1}^{\infty} \frac{1}{3i^2}.$$

Hence, in order to compute $E(Y)$ with precision $\epsilon > 0$, it is sufficient to choose n large enough that $\sum_{i=n+1}^{\infty} (3i^2)^{-1} < \epsilon/2$ and then to compute $E(|X_n|)$ by simulation with accuracy $\epsilon/2$. $\square$

In practice, a random variable Y defined in $([0, 1]^s, \lambda_s)$ (resp., in $([0, 1]^\infty, \lambda_\infty)$) will be said to be *effectively* in L^1 if:

(i) There is a sequence (a_n) decreasing to zero that satisfies

$$\int_{|Y|>n} |Y|\,dP \leq a_n \qquad \forall n \geq 1.$$

Such a sequence is called an *effective bound.*

(ii) For every $k \geq 1$, there is a sequence $(f_{n,k}, n \geq 1)$ of polynomials (resp., cylindrical polynomials) with rational coefficients such that the convergence to zero of

$$\|f_{n,k} - ((Y \wedge k) \vee (-k))\|_{L^1}$$

is proved with an effective bound.

A.4 Integration of Functionals of Stochastic Processes

It is often necessary to deal with random variables that are naturally defined in $([0,1]^\infty, \lambda_\infty)$. For instance, let (X_n) be a Markov chain defined by

$$X_{n+1} = F(X_n, n, U_{n+1}), \qquad X_0 = x,$$

where $F\colon \mathbb{R}^d \times \mathbf{N} \times [0,1]^k \to \mathbb{R}^d$ is given and (U_n) is a Borel sequence in $[0,1]^k$. Actually (X_n) can be defined on the probability space $([0,1]^k, \lambda_k)^{\otimes \mathbf{N}^*}$, where the U_n are the coordinate mappings.

We want to find good double sequences

$$\vec{u}_1 = (u_{11}, \ldots, u_{i1}, \ldots) \in ([0,1]^k)^{\mathbf{N}^*}$$
$$\vdots$$
$$\vec{u}_j = (u_{1j}, \ldots, u_{ij}, \ldots) \in ([0,1]^k)^{\mathbf{N}^*}$$
$$\vdots$$

in such a way that, for $H : (\mathbb{R}^d)^{\mathbf{N}} \to \mathbb{R}$, the average

$$\frac{1}{N} \sum_{j=1}^{N} H(X_\cdot(\vec{u}_j))$$

converges rapidly to $\mathbb{E}[H(X_\cdot)]$.

A.4.1 Exercise

Suppose that (ξ_n) is u.d. in $[0, 1]$, that (V_i) is a Borel sequence in $[0, 1]$, and let $u_{ij} = \{V_i + \xi_j\}$ for $i, j \in \mathbf{N}^*$. Show that the sequence

$$\vec{u}_j = (u_{1j}, \ldots, u_{ij}, \ldots)$$

is not almost surely λ_∞-distributed.

A.4.2 Exercise
Suppose that $u_{ij} \in [0,1]^k$ for i,j in $\mathbf{N}^*$ and let $\mu = \lambda_k^{\otimes \mathbf{N}^*}$. Show that the sequence $(u_j = (u_{1j}, \ldots, u_{ij}, \ldots))_{j \geq 1}$ is μ-distributed iff, for every $q \in \mathbf{N}^*$, the sequence

$$(u_j^q = (u_{1j}, \ldots, u_{qj}))_{j \geq 1}$$

is $\lambda_k^{\otimes q}$-distributed.

A.4.3 Remark
The double sequences that are at present known to yield λ_∞-distributed sequences, for instance, those that are obtained from Halton sequences by putting

$$u_{ij} = \varphi_{p_i}(j) \qquad p_i = i\text{th prime number},$$

are inefficient in practice.

A.4.4 Exercise
Let X_n be a Markov chain with values in $E = \mathbb{R}^k$ and suppose that (X_n) is the sequence of coordinate mappings in the canonical space $E^{\mathbf{N}}$. Let C be a compact subset of E with boundary $\partial C = C \backslash \mathring{C}$ and entrance time $T = \inf\{n \geq 0 : X_n \in C\}$. Suppose that

(i) $\mathbb{P}(T < \infty) = 1$;
(ii) $\mathbb{P}(X_T \in \partial C) = 0$.

Show that X_T is Riemann integrable. [*Hint*: If $X_T(\omega) \in \mathring{C}$, we have $T(\omega) = T(\omega')$ for any ω' close enough to ω, and X_T is continuous at point ω.]

A.4.5 Remark
The notion of a $\mathbb{P}$-Riemann integrable function depends on the topology. If several natural topologies give rise to the same Borel σ-field, such as in the Wiener space setting, the associated classes of $\mathbb{P}$-Riemann integrable functions may be more or less restricted, or even incomparable (cf. Chapter 5).

A.4.6 Discrepancy in Infinite Dimensions By Exercise A.4.2, a sequence $\xi = (\xi_n)_{n\in\mathbf{N}}$ of points in $[0,1]^{\mathbf{N}}$ is $\mathbb{P}$-distributed, where $\mathbb{P} = dx^{\otimes\mathbf{N}}$, if and only if for every k the sequence of its first k coordinates is u.d. on $[0,1]^k$. For every k the discrepancy of the projected sequence $D^{(k)}(\xi, N)$ is well-defined and thus

$$(\xi \text{ is } \mathbb{P}-\text{distributed}) \Leftrightarrow (\forall k \in \mathbf{N}^*, D^{(k)}(\xi, N) \to_{N\to\infty} 0).$$

If α_k is a sequence of positive real numbers such that $\sum \alpha_k = 1$, we can add the equivalence

$$\Leftrightarrow \left(\lim_{N\to\infty} \sum_{k=1}^{\infty} \alpha_k D^{(k)}(\xi, N) = 0 \right).$$

This is valid when $D^{(k)}$ is any of the discrepancies D_∞^* or D_p^*, for $p \in [1, \infty[$. This weighted criterion is, however, not symmetrical with respect to coordinates. We can illustrate this more explicitly as follows.

Let $\mathcal{F}$ be a class of functions from $[0,1]^N$ to $[0,1]$ and let $\xi = (\xi_n)_{n\in\mathbf{N}}$ be a sequence of points in $[0,1]^{\mathbf{N}}$. Consider the criterion

$$D^{\mathcal{F}}(\xi, N) = \sup_{f\in\mathcal{F}} \left| \int f \, d\mathbb{P} - \frac{1}{N} \sum_{n=0}^{N-1} f(\xi_n) \right|.$$

The following result is due to Sobol' [1974].

Proposition A.4.7 Let $X_i, i \in \mathbf{N}$, be the coordinate mappings of $[0,1]^{\mathbf{N}}$. If the class $\mathcal{F}$ contains the functions $f_{ij} = (X_i - X_j)^2, \forall i,j \in \mathbf{N}$, then there does not exist a sequence ξ such that $D^{\mathcal{F}}(\xi, N)$ tends to zero as $N \uparrow \infty$.

Proof Let $\xi = (\xi_n)$ be a sequence of points in $[0,1]^{\mathbf{N}}$. Suppose that there is an N such that $D^{\mathcal{F}}(\xi, N) \le \frac{1}{12}$.

Let M be an integer, and consider the partition of $[0,1]$ into 2^M dyadic intervals I_k of length $|I_k| = 2^{-M}$. For $x \in [0,1]$, let $k(x)$ be the index of the interval I_k such that $x \in I_k$. Each column in the following matrix

$$\begin{array}{cccc} k(X_0(\xi_0)) & k(X_1(\xi_0)) & \dots & k(X_i(\xi_0)) & \dots \\ k(X_0(\xi_1)) & k(X_1(\xi_1)) & \dots & k(X_i(\xi_1)) & \dots \\ \dots & & & & \\ k(X_0(\xi_{N-1})) & k(X_1(\xi_{N-1})) & \dots & k(X_i(\xi_{N-1})) & \dots \end{array}$$

is an array with N components taking integer values between 1 and 2^M. There are thus two identical columns in the matrix, say, columns i and j. Then

$$|X_i(\xi_n) - X_j(\xi_n)| \le 2^{-M} \qquad \forall n = 0, \dots, N-1.$$

Hence

$$\frac{1}{N}\sum_{n=0}^{N-1}(X_i(\xi_n) - X_j(\xi_n))^2 \le 2^{-2M}.$$

Thus the function $f_{ij} = (X_i - X_j)^2$ satisfies

$$\int f_{ij}\, d\mathbb{P} - \frac{1}{N}\sum_{n=0}^{N-1} f_{ij}(\xi_n) \ge \frac{1}{6} - 2^{-2M}.$$

Since $f_{ij} \in \mathscr{F}$ $\forall i, j$ and M is arbitrary, we get

$$\sup_{f \in \mathscr{F}} \left| \int f\, d\mathbb{P} - \frac{1}{N}\sum_{n=0}^{N-1} f(\xi_n) \right| \ge \frac{1}{6},$$

which contradicts the choice of N. Therefore $D^{\mathscr{F}}(\xi, N)$ is bounded away from zero. □

A.5 The Shift Method

The infinite- or high-dimensional case turns out to have its own features that allow us to avoid having to take an average on completely different sample paths.

A.5.1 An Efficient Method

Although the pointwise ergodic theorem has no typical speed of convergence (such as the iterated logarithm law), we give here the method that seems at present to be the most efficient for computing high-dimensional

functionals of stochastic processes: the implementation of this algorithm makes great use of the notion of pointer.

Let us consider the preceding Markov chain

$$X_{n+1} = F(X_n, n, U_{n+1}), \qquad X_0 = x,$$

where (U_n) is the Borel sequence in $[0, 1]$ given by the coordinate mappings of $([0, 1]^\infty, \lambda_\infty)$.

Consider, for instance, the hitting time $T = \inf\{n \geq 1: X_n \in A\}$ and the functional $G(X_T, T)$. The expectation $\mathbb{E}[G(X_T, T)]$ can be computed by Birkhoff's theorem

$$(*) \qquad \mathbb{E}[G(X_T, T)] = \lim_{n\uparrow\infty} \frac{1}{N} \sum_{n=0}^{N-1} G(X_T, T) \circ \theta^n \qquad \text{a.s.}$$

whenever $G(X_T, T) \in L^1$, where θ is the shift operator satisfying

$$\forall \omega \in [0, 1]^\infty \qquad U_n \circ \theta(\omega) = U_{n+1}(\omega), \qquad \forall n \geq 1.$$

The use of pointers can be explained as follows. In order to apply $(*)$ we have to compute $G(X_T, T)$ at successive points $\omega, \theta(\omega), \theta^2(\omega), \ldots$ in $[0, 1]^\infty$. These points are sequences of points in $[0, 1]$:

$$\begin{aligned} \omega &= (U_1(\omega), U_2(\omega), \ldots, U_k(\omega), \ldots) \\ \theta(\omega) &= (U_2(\omega), U_3(\omega), \ldots, U_{k+1}(\omega), \ldots) \\ &\vdots \end{aligned}$$

If, as we suppose here, the stopping time T is finite, we need only compute a finite length of each of these sequences: the sequence ω is computed until $T(\omega)$, which is given by the test

$$\langle \text{test} \rangle \begin{cases} X_k(\omega) \text{ still outside } A \Rightarrow k < T(\omega) \\ X_k(\omega) \in A \text{ for the first time} \Rightarrow k = T(\omega). \end{cases}$$

Suppose we have picked out the sequence $\omega = (U_1, U_2, \ldots)$, and we have put the numbers $U_1, U_2, \ldots$ in pointers indicated in Figure 3.1. The sequence $\theta(\omega) = (U_2, U_3, \ldots)$ is already partially chosen: either it is long enough or it must be lengthened, in which case we have the new scheme (Figure 3.2).

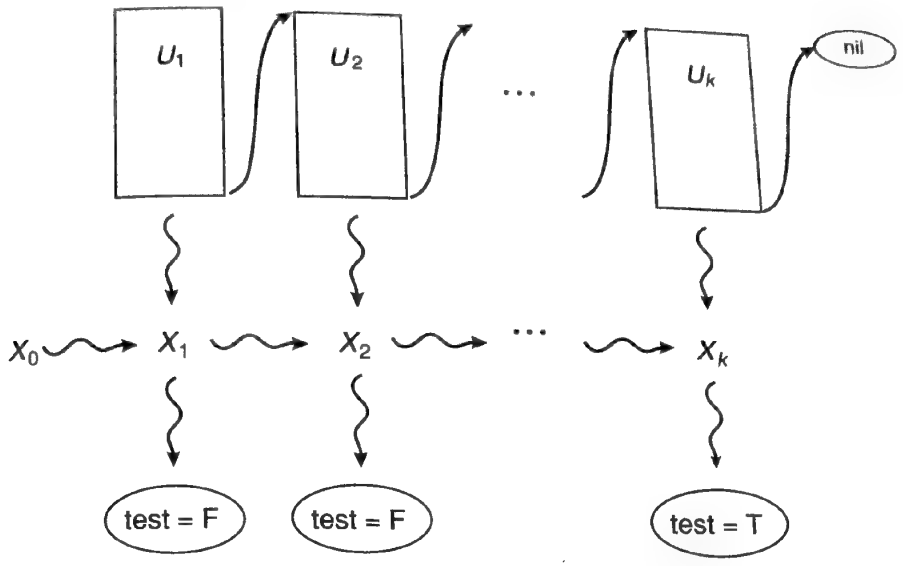


Figure 3.1

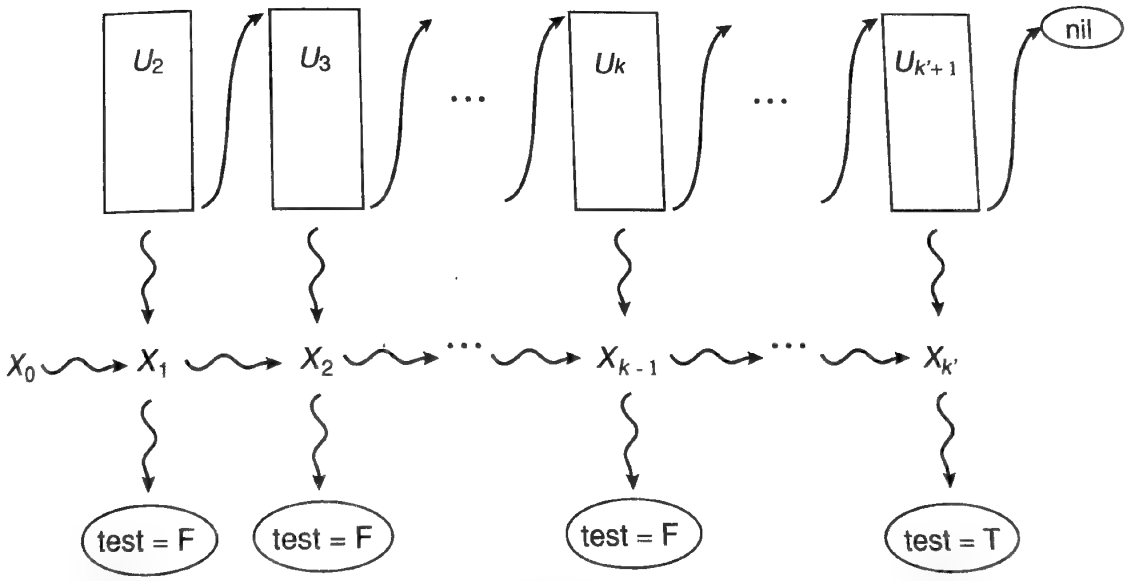


Figure 3.2

In these figures the arrows $\rightsquigarrow$ represent computations to perform, specifically, to compute

$$X_{n+1} = F(X_n, n, U_{n+1}).$$

When we compute $F(x, n, y)$, however, only the variable x (and n) is new and partial computations depending only on y can be stored in the box of U_n in such a way that the remaining computations are easy. This storage must of course be made while we are lengthening the sequence as explained above.

A.5.2 Remarks

(1) We shall see later in this chapter and in Chapter 5 that there are in general many different ways to simulate a stochastic process, based on different representations on the space $([0,1], \mathscr{B}([0,1]), \lambda_1)^{\otimes \mathbf{N}}$. In the case of Brownian motion, this is considered in Section F of Chapter 5. The efficiency of the shift method may depend strongly on the choice of representation (cf. Chapter 5, Section B.4).

(2) The results for Markov chains are of course valid for a large class of stochastic processes or stochastic fields, which, in most cases, can be defined in a natural way on $([0,1], \mathscr{B}([0,1]), \lambda_1)^{\otimes \mathbf{N}}$ after discretization if necessary.

(3) It may happen that the process whose functionals are studied is itself mixing or simply ergodic. In this case, the shift method can be applied to the process itself. This works as soon as a sample path of arbitrary length can be simulated.

(4) The method can be slightly improved by using the two-sided Bernoulli shift. The idea is to simulate two independent copies of the process, one with the classical right to left shift, the other with a left to right shift beginning with the box left by the first one. The benefit is negligible, though, in high dimensions.

A.5.3 Connection with Normal Numbers

For the sake of simplicity, let us replace $[0,1]^{\mathbf{N}}$ by $\{0,1\}^{\mathbf{N}}$. The best points for integration by the shift are fast normal numbers. Recall that a real number x in [0,1], or a sequence of binary digits, is called *normal* if the sequence $\theta^n(x)$ is μ-distributed on $\{0,1\}^{\mathbf{N}}$ equipped with the measure $\mu = (\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1)^{\otimes \mathbf{N}}$. Explicit normal numbers are known, for instance, the Champernowne sequence. Other normal sequences are known, which are generalizations of Champernowne's (e.g., see Rauzy [1976, 1981] or Dumont and Thomas [1986]). Some of these constructions make use of (s,m)-de Bruijn sequences, which are periodic mappings u from $\mathbf{N}$ into $\{0,\ldots,m-1\}$ of period m^s such that the m^s strings of length s

$$(u(n),\ldots,u(n+s+1)), \qquad n = 0,1,\ldots,m^s-1,$$

are all different. Algorithms are known for generating infinite de Bruijn sequences (see Ivanyi [1988/89] and its references), that is, infinite sequences whose initial sections of length m^s, considered as loops, are (s,m)-de Bruijn sequences for every s.

The efficiency of these sequences is difficult to evaluate, and is largely an open question. One of the main difficulties is to summarize the perfor-

mance of a sequence by a single scale, when there are naturally infinitely many discrepancies, one for each finite product space.

Flajolet et al. [1988] give an asymptotic result on the space $\{0,1\}^{\mathbf{N}}$, which shows that, in a certain sense, random sequences are not far from being the best.

A.6 Rate of Convergence for the Shift Method

In contrast to the case of the law of large numbers, there is no standard speed of convergence that holds for every function in L^2 in the setting of the pointwise ergodic theorem. Nevertheless, the presently known functions that integrate very slowly or very quickly by the shift method are constructed by suitable applications of the Rohlin–Halmos lemma. For commonly met functions a law of iterated logarithm holds with a new coefficient different from the standard deviation. We quote here some results that are derived from the theorems of Gàl and Koksma [Ben Alaya, 1992] and Heyde and Scott [Bouleau (6)].

A.6.1 Slow Functions and Fast Functions

It is known [Halasz, 1976; Krengel, 1978, 1985] that for every ergodic endomorphism θ of a Polish space $(\Omega, \mathscr{A}, \mathbb{P})$, and for every sequence (α_n) such that $\alpha_n \to 0$, there is an f in L^2 such that

$$\overline{\lim_{N\uparrow\infty}} \left| \frac{1}{N} \sum_{n=0}^{N-1} f \circ \theta^N - \mathbb{E}f \right| \Big/ \alpha_N = +\infty \qquad \text{a.s.}$$

The existence of such a "slow" function is proved by using the Rohlin–Halmos lemma [Cornfeld, Fomin, and Sinai, 1982]; similarly, for every $\epsilon > 0$, the lemma furnishes a non constant "fast" function $f \in L^2$ for which

$$\frac{1}{N} \sum_{n=0}^{N-1} f \circ \theta^N - \mathbb{E}f = o\left(\frac{1}{N^{1-\epsilon}} \right) \qquad \text{a.s.}$$

A.6.2 Functions Adapted to a Stopping Time

Let $(\Omega, \mathscr{A}, P) = ([0,1], \mathscr{B}([0,1]), \lambda_1)^{\otimes \mathbf{N}}$. Let $(X_n)_{n\geq 0}$ be the coordinate mappings on Ω, and define the shift θ on Ω by

$$X_n \circ \theta = X_{n+1} \qquad \forall n \in \mathbf{N}.$$

For $n \geq 0$, put

$$\mathscr{F}_0^n = \sigma(X_0, \ldots, X_n),$$

and let T be a stopping time with respect to the filtration $\mathscr{F} = (\mathscr{F}_0^n)_{n\geq 0}$. We are interested in the $\mathscr{F}_0^T$-measurable random variables. Such functions are very commonly used in simulation problems, and they turn out to be the broadest class of random variables that can be simulated (cf. Chapter 5).

For F in $L^2(\Omega, \mathscr{A}, \mathbb{P})$ with $\mathbb{E}(F) = 0$, put

$$s(F) = (\mathbb{E}(F^2) + 2\sum_{k=1}^{\infty} \mathbb{E}(F \circ \theta^k . F))^{1/2}.$$

The following statements are proved in Ben Alaya [1992].

Theorem A.6.3 Let T be a $\mathscr{F}$-stopping time and assume T has a finite moment of order $p > 2$. If $F \in L^2, \mathbb{E}(F) = 0$ and F is $\mathscr{F}_0^T$-measurable, then:

(i)

$$|\mathbb{E}(F \circ \theta^k . F)| \leq \frac{\mathbb{E}(F^2)(\mathbb{E}(T^p))^{1/2}}{k^{p/2}} \qquad k \geq 1;$$

(ii) if $s(F) > 0$, then

$$\frac{1}{s(F)\sqrt{N}} \sum_{n=0}^{N-1} F \circ \theta^n \xrightarrow{\mathscr{L}} N(0, 1);$$

(iii) for every $\epsilon > 0$,

$$\frac{1}{N} \sum_{n=0}^{N-1} F \circ \theta^n = o(N^{-1/2}(\log N)^{3/2+\epsilon}) \qquad \text{a.s.}$$

From the standpoint of computer efficiency, it makes little difference whether we shift forward or backward. However, theoretical results are more conveniently stated with the operator $\tau = \theta^{-1}$. We consider a prob-

ability space $(E, \mathscr{E}, \mu)$ and put

$$(\Omega, \mathscr{A}, \mathbb{P}) = (E, \mathscr{E}, \mu)^{\otimes \mathbf{Z}};$$

the coordinate mappings are denoted by X_n $(n \in \mathbf{Z})$ and τ is defined on Ω by

$$X_n \circ \tau = X_{n-1} \qquad \forall n \in \mathbf{Z}.$$

Put

$$\begin{aligned}
\mathscr{F}_m^n &= \sigma(X_m, X_{m+1}, \ldots, X_n) \qquad m \leq n \\
\mathscr{F}_{-\infty}^n &= \sigma(X_k, k \leq n) \\
\mathscr{F}_m^{+\infty} &= \sigma(X_k, k \geq m) \\
\mathscr{F}_{-\infty}^{+\infty} &= \sigma(X_k, k \in \mathbf{Z})
\end{aligned}$$

Let T be the operator defined on $L^1(\Omega, \mathscr{F}_0^{+\infty}, \mathbb{P})$ by

$$Tf = \mathbb{E}[f | \mathscr{F}_1^{+\infty}] \circ \tau.$$

Lemma A.6.4 For $f \in L^2(\Omega, \mathscr{F}_0^{+\infty}, \mathbb{P})$ with $\mathbb{E}(f) = 0$, the following conditions are equivalent:

(i) $\sum_{n=0}^{N} T^n f$ remains bounded in L^2.
(ii) $\sum_{n=0}^{N} T^n f$ converges weakly in L^2 as $N \uparrow \infty$.
(iii) $\sum_{n=0}^{N} T^n f$ converges in L^2 as $N \uparrow \infty$.
(iv) There exists $g \in L^2(\Omega, \mathscr{F}_0^{+\infty}, \mathbb{P})$ such that $f = (I - T)g$.

We shall say that a function $f \in L^2(\Omega, \mathscr{F}_0^{+\infty}, \mathbb{P})$ belongs to the Gordin class (which we denote by $\mathscr{G}$) if $f - \mathbb{E}(f)$ satisfies the conditions of the above lemma.

Lemma A.6.5 The Gordin class $\mathscr{G}$ is the class of functions f in $L^2(\Omega, \mathscr{F}_0^{+\infty}, \mathbb{P})$ which admit a decomposition

$$(*) \qquad f - \mathbb{E}(f) = \tilde{g} + h \circ \tau^{-1} - h$$

where $\tilde{g}, h \in L^2(\mathscr{F}_0^{+\infty})$ with $\mathbb{E}[\tilde{g} | \mathscr{F}_1^{+\infty}] = 0$ and $\mathbb{E}(h) = 0$. Such a decomposition, if it exists, is unique.

These lemmas are easily proved [Bouleau (6)]. By using a theorem of Heyde and Scott [1973], one can deduce a law of the iterated logarithm.

Theorem A.6.6 Let $f \in L^2(\Omega, \mathscr{F}_0^{+\infty}, \mathbb{P})$ be in the Gordin class, and let $\tilde{g}, h$ be the elements of its decomposition (*). Put

$$S_N = \sum_{n=0}^{N} (f - \mathbb{E}(f)) \circ \tau^n.$$

Then

(i) $\lim_{N\uparrow\infty}(1/\sqrt{N})\|S_N\|_2 = \|\tilde{g}\|_2$;
(ii) $\overline{\lim}_{N\uparrow\infty}(|S_N|/(2N \log\log N)^{1/2}) = \|\tilde{g}\|_2$.

We refer to Bouleau [(6)] for a proof and for sufficient conditions ensuring membership of the Gordin class $\mathscr{G}$.

In the case of $[0,1]^s$, the membership of the Gordin class can be expressed in terms of local smoothness. Consider the following automorphism of $[0,1]^s \times [0,1]^s$:

$$((x_1, \ldots, x_s), (y_1, \ldots, y_s)) \overset{\tau}{\rightsquigarrow} ((\{2x_1\}, \ldots, \{2x_s\}), (\tfrac{1}{2}([2x_1] + y_1), \ldots, \tfrac{1}{2}([2x_s] + y_s)).$$

This automorphism is sometimes referred to as the "baker's transformation" and it is easily seen that it corresponds to the bilateral Bernoulli shift on the binary expansion of real numbers.

> [La transformation du boulanger] Il ne s'agit pas d'une transformation dynamique hamiltonienne, mais, étant donné qu'elle préserve la mesure, on peut s'en servir pour illustrer de nombreux aspects de flots hamiltoniens.[1] [Ilya Prigogine, 1980]

Proposition A.6.7 Let f be in $L^2([0,1]^s, \lambda_s)$ and have Fourier expansion

$$f(y) = \sum_{m \in \mathbf{Z}^s} a_m e^{2i\pi\langle m, y\rangle}.$$

[1] [The baker's transform] is not a question of a transformation in hamiltonian dynamics, but since it is measure-preserving it can be used to illustrate many aspects of hamiltonian flows.

(i) If there exists a sequence (c_n), $c_n \geq 0$, $\sum_{n=0}^{\infty} c_n < \infty$ such that

$$|a_{2^n m}| \leq c_n |a_m| \qquad \forall m \in \mathbf{Z}^s, \quad \forall n \geq 0$$

then $f \in \mathscr{G}$ and the function $\tilde{g}$ in its decomposition (*) satisfies

$$\|\tilde{g}\|_2 \leq \|f - \mathbb{E}(f)\|_2 \sum_{n=0}^{\infty} c_n.$$

(ii) If there exists $\alpha > 0$, which satisfies

$$\sum_{p \in \mathbf{Z}^s} |a_p|^2 \left(\sum_{i=1}^{s} p_i^2 \right)^{\alpha} < \infty,$$

then $f \in \mathscr{G}$ and there exists a constant $c > 0$, independent of f, such that

$$\|\tilde{g}\|_2 \leq c \left(\sum_{p \in \mathbf{Z}^s} |a_p|^2 \left(\sum_{i=1}^{s} p_i^2 \right)^{\alpha} \right)^{1/2}.$$

In the case of the infinite-dimensional torus $[0,1]^{\mathbf{N}}$, membership of the Gordin class is clearly related to the size of the derivatives of f with respect to the coordinates with large index. A Dirichlet form method (cf. Section A.8 below) yields the following result.

Proposition A.6.8 Put $(\Omega, \mathscr{A}, \mathbb{P}) = ([0,1]^s, \mathscr{B}([0,1]^s), \lambda_s)^{\otimes \mathbf{Z}}$. Let $f \in L^2(\Omega, \mathscr{F}_0^{+\infty}, \mathbb{P})$ have the following property: for every $n \in \mathbf{N}$, the map

$$[0,1]^s \ni x_n \rightsquigarrow f(x_0, \ldots, x_n, \ldots) \in \mathbb{R}$$

has $(dx_0 \cdots dx_{n-1} dx_{n+1} \cdots)$-almost everywhere a derivative in the sense of distributions belonging to $L^2(dx_n)$. If

$$\sum_{k=0}^{\infty} \left(\sum_{i=k}^{\infty} \mathbb{E}[|\nabla_i f|^2] \right)^{1/2} < \infty$$

or if

$$\sum_{i=2}^{\infty} i^2(\log^{\alpha} i)\mathbb{E}[|\nabla_i f|^2] < \infty$$

for some $\alpha > 1$, then $f \in \mathscr{G}$.

See Bouleau [(6)] for a proof.

A.7 On the Shift Method in Finite Dimension

The shift method can also be useful in finite dimensions, and a bit more can be said in this case.

A.7.1 Case of the Shift Method with a Borel Sequence

Consider $f : [0,1]^p \to \mathbb{R}$ seen as a function from $[0,1]^{\mathbf{Z}}$ to $\mathbb{R}$ that depends only on p coordinates. As we have seen, if $f \in L^2(dx)$ then f belongs to the Gordin class, and this is true for both shifts θ or $\theta^{-1} = \tau$.

We introduce the notation

$$\widetilde{\mathrm{Var}}\, f = \|\tilde{g}\|_2^2,$$

where

$$f - \mathbb{E}f = \tilde{g} + h \circ \tau^{-1} - h$$

is the Gordin decomposition of f (with respect to τ). $\widetilde{\mathrm{Var}}\, f$ can easily be computed by computing $\|S_N\|_2^2$ and looking at $\lim_{N\uparrow\infty}(1/N)\|S_N\|_2^2$; we obtain

$$\widetilde{\mathrm{Var}}\, f = \mathrm{Var}\, f + 2\sum_{k=1}^{p-1} \mathrm{Cov}\,(f, f \circ \tau^k),$$

which of course is equal to

$$\mathrm{Var}\, f + 2\sum_{k=1}^{p-1} \mathrm{Cov}\,(f, f \circ \theta^k).$$

Thus $\widetilde{\mathrm{Var}}\, f$ does not depend on the direction of the shift. Note that $\widetilde{\mathrm{Var}}\, f$ can be either greater or less than the variance $\mathrm{Var}\, f$.

Now, if f is modified by a permutation π of its coordinates

$$f_\pi(x_1, \ldots, x_p) = f(x_{\pi(1)}, \ldots, x_{\pi(p)}),$$

then the quantities $\widetilde{\mathrm{Var}}\, f_\pi$ and $\widetilde{\mathrm{Var}}\, f$ are in general different. Furthermore, if, instead of applying the simple shift, we apply the double shift θ^2 or the shift θ^k, then the quantity $\widetilde{\mathrm{Var}}\, f$ will also change.

Thus we have $p.p!$ methods for computing $\mathbb{E}f$ (the shift θ^p is the classical Monte Carlo method), which in general have different coefficients in the law of the iterated logarithm.

Finally we quote the result that, if f is bounded, the shift has an exponential rate of convergence in the sense of Chernov theorem: $\forall\, \varepsilon > 0, \exists C < +\infty$ and $\gamma > 0$ such that

$$\mathbb{P}\left[\frac{1}{N}\sum_{n=0}^{N-1} f \circ \theta^n > \mathbb{E}f + \varepsilon\right] \leq Ce^{-\gamma N}.$$

A.7.2 Case of the Shift with p-Uniformly Distributed Sequences

A sequence $\xi = (\xi_n)$ of points in $[0, 1]$ is said to be p-uniformly distributed if the sequence

$$\tilde{\xi} = (\xi_n, \xi_{n+1}, \ldots, \xi_{n+p-1})_{n \geq 1}$$

is uniformly distributed on $[0, 1]^p$.

Integrating a Riemann integrable function $f : [0, 1]^p \to \mathbb{R}$ by using a p-uniformly distributed sequence in the formula

$$\mathbb{E}f = \lim_N \frac{1}{N}\sum_{n=1}^{N} f(\xi_n, \xi_{n+1}, \ldots, \xi_{n+p-1})$$

clearly presents the same advantages in implementation, namely, being able to store partial calculations, as the shift method presents with a Borel sequence. The rate of convergence depends here on the discrepancy of the sequence $\tilde{\xi}$.

Such p-u.d. sequences can be constructed by the following Van der Corput criterion.

Lemma A.7.3 Suppose that $u = (u_n)$ is a sequence of real numbers such that $\forall p \geq 1$ the sequence $u_{n+p} - u_n$ is uniformly distributed modulo 1. Then the sequence (u_n) is uniformly distributed modulo 1.

This result can be proved either by expanding the square

$$\left| \frac{1}{N} \sum_{n<N} e^{2i\pi u_n} \right|^2$$

and estimating the resulting rectangle terms [Rauzy, 1976] or by using correlation functions and Bochner's theorem [Rauzy, 1976; Bass, 1984]. We get the following result.

Proposition A.7.4 Let $P(x) = a_0 + \cdots + a_p x^p$ be a polynomial of degree $p \geq 1$ with a_p irrational (Weyl polynomial). Then the sequence $P(n)$ is p-uniformly distributed modulo 1.

Proof By Weyl's criterion (Chapter 2, Proposition A.5.3) it suffices to show that for every $m = (m_1, \ldots, m_p) \in \mathbf{Z}^p, m \neq (0, \ldots, 0)$, the sequence

$$u_n = m_1 P(n+1) + \cdots + m_p P(n+p)$$

is uniformly distributed modulo 1. We have

$$u_n = \left(\sum_{i=1}^{p} m_i \right) a_p n^p + \sum_{i=1}^{p} m_i (C_p^1 i a_p + a_{p-1}) n^{p-1} + \cdots$$
$$\cdots + \sum_{i=1}^{p} m_i [C_p^k i^k a_p + C_{p-1}^{k-1} i^{k-1} a_{p-1} + \cdots + a_{p-k}] n^{p-k} + \cdots$$
$$\cdots + \sum_{i=1}^{p} m_i [i^p a_p + i^{p-1} a_{p-1} + \cdots + a_0].$$

The numbers

$$\sum_{i=1}^{p} m_i i^k \qquad k = 0, 1, \ldots p-1$$

cannot all vanish. u_n is a polynomial in n whose term of highest degree is of the form

$$\sum_{i=1}^{p} m_i i^k C_p^k a_p,$$

where k is the smallest integer for which $\sum_{i=1}^{p} m_i i^k \neq 0$. It is enough to prove that when Q is a polynomial of degree ≥ 1, such that the term of highest degree in Q has irrational coefficient, then $Q(n)$ is u.d. modulo 1.

If Q has degree 1, then it is an irrational translation on the torus, which was dealt with in Chapter 2, Section C.3. If Q has degree $d > 1$, then $\forall p \geq 1, Q(x+p) - Q(x)$ is a polynomial of degree $d-1$ whose term of highest degree has an irrational coefficient; by the inductive hypothesis, $Q(n+p) - Q(n)$ is u.d. modulo 1, and so $Q(n)$ is u.d. modulo 1 by the lemma. □

A.8 Complement: Quasi-every Sequence Is Shift Uniformly Distributed

Consider the probability space $([0,1], \mathscr{B}([0,1]), \lambda_1)^{\otimes \mathbf{N}}$. A sequence $\omega \in [0,1]^{\mathbf{N}}$ is said to be θ-u.d. (also "normal" or "completely u.d.") if the probability measures

$$\frac{1}{N} \sum_{n=0}^{N-1} \delta_{\theta^n(\omega)},$$

where θ is the shift operator, converge narrowly to $\lambda_1^{\otimes \mathbf{N}}$. It follows easily from the pointwise ergodic theorem that $\lambda_1^{\otimes \mathbf{N}}$-almost every sequence is θ-u.d. By analogy with the case $(\{0,1\}, \mathscr{B}(\{0,1\}), \{\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1\})^{\otimes \mathbf{N}}$, it is nevertheless suspected that there are more θ-u.d. sequences than "random" sequences. Actually, as we will now see, the set of non θ-u.d. sequences has null capacity for a large class of capacities on $[0,1]^{\mathbf{N}}$.

A.8.1 Dirichlet Structures

Let $(\Omega, \mathscr{F}, m)$ be a probability space. We impose extra structure on it by adding a *Dirichlet form* $\mathscr{E}$ with domain $\mathbb{D}$, that is, a symmetric nonnegative definite bilinear form $f, g \rightsquigarrow \mathscr{E}(f,g)$ defined on a dense subspace $\mathbb{D}$ of $L^2(m)$, which satisfies the following two conditions:

(i) $\mathscr{E}$ is closed, that is, $\mathbb{D}$ is complete with respect to the norm

$$\|f\|_{\mathbb{D}} = (\|f\|_{L^2}^2 + \mathscr{E}(f,f))^{1/2};$$

(ii) if $f \in \mathbb{D}$, then $f \wedge 1 \in \mathbb{D}$ and

$$\mathscr{E}(f \wedge 1, f \wedge 1) \leq \mathscr{E}(f, f).$$

We obtain what is called a *Dirichlet structure* $(\Omega, \mathscr{F}, m, \mathscr{E}, \mathbb{D})$. Let F be a contraction from $\mathbb{R}$ to $\mathbb{R}$ (i.e., $F(0) = 0$ and $|F(x) - F(y)| \leq |x - y|$ for every x, y in $\mathbb{R}$). It can be proved [Fukushima, 1980; Bouleau and Hirsch, 1991] that if $f \in \mathbb{D}$, then

$$F \circ f \in \mathbb{D} \qquad \text{and} \qquad \mathscr{E}(F \circ f, F \circ f) \leq \mathscr{E}(f, f).$$

We shall in addition assume that

(H) $1 \in \mathbb{D}$ and $\mathscr{E}(1, 1) = 0$

(T) Ω is a topological space with Borel σ-field $\mathscr{F}$.

A.8.2 Examples

(1) Let $\Omega = [0, 1], \mathscr{F} = \mathscr{B}([0, 1]), m = \lambda_1$

$$\mathbb{D} = H^1 = \left\{ u \in L^2(\lambda_1), u \text{ is absolutely continuous with derivative } u' \text{ and } \int_0^1 (u'(x))^2 \, dx < \infty \right\}$$

$$\mathscr{E}(u, v) = \int_0^1 u'(x) v'(x) \, dx.$$

Proof of (i): Let (u_n) be a Cauchy sequence with respect to the norm $\|.\|_{\mathbb{D}}$. Then there exist u in L^2 and f in L^2 such that $u_n \to u$ in L^2 and $u_n' \to f$ in L^2. It follows from the inequality

$$|u(y) - u(x)|^2 = \left| \int_x^y u'(t) \, dt \right|^2 \leq |y - x| \int_0^1 (u'(t))^2 dt$$

and the boundedness of the sequence $\|u_n'\|_2$ that the sequence (u_n) is equicontinuous. Thus there is a subsequence (u_{n_k}) that converges uniformly to a continuous function $\tilde{u} = u$ a.e. Let $\varphi \in \mathscr{D}$ vanish outside $]0, 1[$. Then

$$\int_0^1 f\varphi\,dx = \lim_{n\uparrow\infty}\int_0^1 u_n'\varphi\,dx = -\lim_{n\uparrow\infty}\int_0^1 u_n\varphi'\,dx = -\int_0^1 \tilde{u}\varphi'\,dx.$$

Thus $\tilde{u}$ is absolutely continuous with derivative f, $u_n \to \tilde{u}$ in $\mathbb{D}$ and so the form $\mathscr{E}$ is closed.

The proof of (ii) is left as an exercise.

(2) Product spaces. Let $([0,1], \mathscr{B}([0,1]), \mu, \mathbf{e}, \mathbf{d})$ be a Dirichlet structure. We shall define a new Dirichlet structure

$$(\Omega, \mathscr{F}, m, \mathscr{E}, \mathbb{D}) = ([0,1], \mathscr{B}([0,1]), \mu, \mathbf{e}, \mathbf{d})^{\otimes \mathbf{N}}$$

by putting

$$(\Omega, \mathscr{F}, m) = ([0,1], \mathscr{B}([0,1]), \mu)^{\otimes \mathbf{N}}$$

and $\mathbb{D}$ to be the set of functions $F : [0,1]^{\mathbf{N}} \to \mathbb{R}$ such that $F \in L^2(m)$, for every $n \in \mathbf{N}$ and for almost all fixed $\omega_0, \omega_1, \ldots, \omega_{n-1}, \omega_{n+1}, \ldots$, the function

$$x \rightsquigarrow F(\omega_0, \ldots, \omega_{n-1}, x, \omega_{n+1}, \ldots)$$

belongs to $\mathbf{d}$ and

$$\int \sum_n \mathbf{e}_n(F,F)\,dm < \infty,$$

where $\mathbf{e}_n$ is the form $\mathbf{e}$ acting on the nth coordinate. Finally, for $F \in \mathbb{D}$, put

$$\mathscr{E}(F,F) = \mathbb{E}\left[\sum_{n=0}^{\infty} \mathbf{e}_n(F,F)\right] \quad \left(= \int \sum_{n=0}^{\infty} \mathbf{e}_n(F,F)\,dm\right).$$

Proposition A.8.3 The form $(\mathscr{E}, \mathbb{D})$ is closed.

Proof Let (f_n) be a Cauchy sequence in $\mathbb{D}$ such that $\|f_n - f\|_2 \to 0$, where $f \in L^2(m)$. There is a subsequence (f_{n_k}) such that

$$\sum_k \|f_{n_k} - f\|_2^2 < \infty$$

and

$$\sum_k (\mathscr{E}(f_{n_{k+1}} - f_{n_k}, f_{n_{k+1}} - f_{n_k}))^{1/2} < +\infty.$$

It follows that for any $i \in \mathbf{N}$, for almost all $\omega_0, \ldots, \omega_{i-1}, \omega_{i+1}, \ldots$, we have

(a) $\sum_k \int (f_{n_k}(\omega_0, \ldots, \omega_{i-1}, x, \omega_{i+1} \ldots) - f(\omega_0, \ldots, \omega_{i-1}, x, \omega_{i+1} \ldots))^2 d\mu(x) < \infty$;

(b) $\sum_k (\mathbf{e}_i(f_{n_{k+1}} - f_{n_k}, f_{n_{k+1}} - f_{n_k}))^{1/2} < \infty$.

Assertion (b) follows from the inequalities

$$\mathbb{E}[(\mathbf{e}_i(F,F))^{1/2}] \le (\mathbb{E}[\mathbf{e}_i(F,F)])^{1/2} \le (\mathscr{E}(F,F))^{1/2}.$$

Since the form $(\mathbf{e}, \mathbf{d})$ is closed, we deduce that for almost all $\omega_0, \ldots, \omega_{i-1}, \omega_{i+1}, \ldots, f(\omega_0, \ldots, \omega_{i-1}, ., \omega_{i+1} \ldots) \in \mathbf{d}$ and $\mathbf{e}_i(f_{n_k} - f, f_{n_k} - f) \to 0$. Therefore by Fatou's lemma

$$\mathbb{E}\left[\sum_{i=0}^{\infty} \mathbf{e}_i(f,f)\right] = \mathbb{E}\left[\sum_{i=0}^{\infty} \lim_k \mathbf{e}_i(f_{n_k}, f_{n_k})\right] \le \varliminf_k \mathscr{E}(f_{n_k}, f_{n_k}).$$

Since (f_n) is a Cauchy sequence in $\mathbb{D}$, the last term is finite and f belongs to $\mathbb{D}$. Another application of Fatou's lemma shows that

$$\begin{aligned}\mathbb{E}\sum_{i=0}^{\infty} \mathbf{e}_i(f - f_n, f - f_n) &= \mathbb{E}\sum_{i=0}^{\infty} \lim_k \mathbf{e}_i(f_{n_k} - f_n, f_{n_k} - f_n) \\ &\le \varliminf_k \mathbb{E}\sum_{i=0}^{\infty} \mathbf{e}_i(f_{n_k} - f_n, f_{n_k} - f_n) \\ &= \varliminf_k \mathscr{E}(f_{n_k} - f_n, f_{n_k} - f_n),\end{aligned}$$

which can be made $\leq \varepsilon$ for large enough k and n, because f_n is a Cauchy sequence for $\|.\|_{\mathbb{D}}$. Thus $f_n \to f$ in $\mathbb{D}$. □

A.8.4 Capacities

Let $(\Omega, \mathscr{F}, m, \mathscr{E}, \mathbb{D})$ be a Dirichlet structure. If G is an open set in Ω, we define

$$c(G) = \inf\{\|u\|_{\mathbb{D}}^2; u \in \mathbb{D} \text{ and } u \geq 1_G\}.$$

If A is any subset of Ω, the capacity of A is defined by

$$c(A) = \inf\{c(G) : G \text{ open}, G \supset A\}.$$

Lemma A.8.5 Let $u, v \in \mathbb{D}$. Then $u \vee v, u \wedge v \in \mathbb{D}$ and

$$\|u \vee v\|_{\mathbb{D}}^2 + \|u \wedge v\|_{\mathbb{D}}^2 \leq \|u\|_{\mathbb{D}}^2 + \|v\|_{\mathbb{D}}^2.$$

Proof We have

$$u \vee v = \tfrac{1}{2}((u+v) + |u-v|) \in \mathbb{D}$$
$$u \wedge v = \tfrac{1}{2}((u+v) - |u-v|) \in \mathbb{D}$$

and

$$\begin{aligned}\|u \vee v\|_{\mathbb{D}}^2 + \|u \wedge v\|_{\mathbb{D}}^2 &= \tfrac{1}{2}(\|u+v\|_{\mathbb{D}}^2 + \||u-v|\|_{\mathbb{D}}^2) \\ &\leq \tfrac{1}{2}(\|u+v\|_{\mathbb{D}}^2 + \|u-v\|_{\mathbb{D}}^2) \\ &= \|u\|_{\mathbb{D}}^2 + \|v\|_{\mathbb{D}}^2.\end{aligned}$$

□

Proposition A.8.6

(i) $\forall A \in \mathscr{F}, m(A) \leq c(A) \leq 1$,
(ii) $\forall A_1, A_2 \subset \Omega, \quad c(A_1 \cup A_2) + c(A_1 \cap A_2) \leq c(A_1) + c(A_2)$,
(iii) $\forall A_n \subset \Omega, A_n \uparrow, \quad c(\cup_n A_n) = \lim_{N \uparrow \infty} c(A_n)$,
(iv) $\forall A_n \subset \Omega \quad c(\cup_n A_n) \leq \sum_{n=0}^{\infty} c(A_n)$.

Proof Assertion (i) follows directly from the definition. Assertion (ii) follows easily from the lemma and the definition. For (iii), we remark that if G is an open set and e_G is the orthogonal projection of 0 on the closed

convex set $\{u \in \mathbb{D}; u \geq 1_G \text{ a.e.}\}$ in $\mathbb{D}$, then $e_G \geq 1_G$ and $\|e_G\|^2_{\mathbb{D}} = c(G)$. If (G_n) is an increasing sequence of open sets, then, by the properties of projections in Hilbert space, we have $e_{G_n} \to e_{\cup_n G_n}$, and thus $c(\sup_n G_n) = \lim_n \uparrow c(G_n)$. Finally, if (A_n) is an increasing sequence of subsets, (iii) is obtained by using (ii) and approximating each A_n with an open subset G_n in the sense of capacity. Assertion (iv) is a straightforward consequence of (ii) and (iii). □

Proposition A.8.7 Let f be a continuous function in $\mathbb{D}$. Then, for any $\alpha > 0$,

$$c\{|f| > \alpha\} \leq \frac{\|f\|^2_{\mathbb{D}}}{\alpha^2}.$$

Proof The set $G = \{|f| > \alpha\}$ is open, $|f| \in \mathbb{D}$ and $\alpha^{-1}|f| \geq 1_G$. □

A set of capacity 0 is said to be a *c-null set*. Relations or convergences valid outside a c-null set are said to be valid *quasi-everywhere* (q.e.).

Proposition A.8.8 Let $(\Omega, \mathscr{F}, m, \mathscr{E}, \mathbb{D})$ be a Dirichlet structure that satisfies (H) and (T). Let (X_n) be a sequence in $\mathscr{C}(\Omega) \cap \mathbb{D}$ such that

- $\mathbb{E}(X_n) = a \qquad \forall n \geq 0$
- $(X_n - a)$ is an orthogonal sequence in $\mathbb{D}$
- $\sup_n \|X_n\|_{\mathbb{D}} < \infty$.

Then

$$\frac{1}{N}\sum_{n=0}^{N-1} X_n \to a$$

in $\mathbb{D}$ and quasi-everywhere.

Proof Put $K = 2(\sup_n \|X_n\|^2_{\mathbb{D}} + a^2) < \infty$.

(a) Since

$$\left\| \frac{1}{N}\sum_{n=0}^{N-1}(X_n - a) \right\|^2_{\mathbb{D}} = \frac{1}{N^2}\sum_{n=0}^{N-1} \|X_n - a\|^2_{\mathbb{D}} \leq \frac{K}{N},$$

convergence in $\mathbb{D}$ is clear.

(b) We claim that

$$\frac{1}{m^2}\sum_{n=0}^{m^2-1}(X_n - a)$$

converges q.e. to 0. Indeed, let

$$S_N = \sum_{n=0}^{N-1}(X_n - a), \qquad N \geq 1$$

$$A_M^{\epsilon} = \bigcup_{m \geq M} \left\{\left|\frac{S_{m^2}}{m^2}\right| > \epsilon\right\} \qquad \epsilon > 0, M \geq 1.$$

We have

$$c(A_M^{\epsilon}) \leq \epsilon^{-2} \sum_{m \geq M} \left\|\frac{S_{m^2}}{m^2}\right\|_{\mathbb{D}}^2$$

$$\leq K\epsilon^{-2} \sum_{m \geq M} \frac{1}{m^2}.$$

Putting

$$A_{\infty}^{\epsilon} = \left\{\lim_m \sup \left|\frac{S_{m^2}}{m^2}\right| > \epsilon\right\} = \bigcap_{M \geq 1} A_M^{\epsilon},$$

we have

$$c(A_{\infty}^{\epsilon}) \leq c(A_M^{\epsilon}) \qquad \text{for any } M \geq 1,$$

and thus $c(A_{\infty}^{\epsilon}) = 0$. If

$$A = \left\{\lim_m \sup \left|\frac{S_{m^2}}{m^2}\right| > 0\right\} = \bigcup_{p \geq 1} A_{\infty}^{1/p},$$

it follows from the subadditivity of c that $c(A) = 0$.

(c) For $n \geq 1$, put

$$m(n) = \sup\{m : m^2 \leq n\}.$$

It remains to show that

$$\left|\frac{1}{n}S_n - \frac{1}{m(n)^2}S_{m(n)^2}\right| \to 0 \qquad \text{q.e.}$$

for which it is enough to show that

$$\frac{1}{n}|S_n - S_{m(n)^2}| \to 0 \qquad \text{q.e.}$$

For $\epsilon > 0$ and $N \geq 1$, put

$$B_N^\epsilon = \bigcup_{n \geq N} \left\{\frac{1}{n}|S_n - S_{m(n)^2}| > \epsilon\right\}.$$

Then

$$c(B_N^\epsilon) \leq \epsilon^{-2} \sum_{n \geq N} \frac{1}{n^2}((n - m(n)^2)K$$

$$\leq \epsilon^{-2} K \sum_{n \geq N} \frac{2\sqrt{n}}{n^2}$$

and the argument used in (b) yields

$$c\left(\limsup_n \frac{1}{n}|S_n - S_{m(n)^2}| > 0\right) = 0. \qquad \square$$

We return to the shift operator θ on $[0,1]^{\mathbf{N}}$.

Proposition A.8.9 Let $(\mathbf{e}, \mathbf{d})$ be a Dirichlet form on $([0,1], \mathscr{B}([0,1]), \lambda_1)$ such that $\mathscr{C}^1([0,1]) \subset \mathbf{d}$. In the Dirichlet structure $([0,1], \mathscr{B}([0,1]), \lambda_1, \mathbf{e}, \mathbf{d})^{\otimes \mathbf{N}}$, the set of sequences $\omega \in [0,1]^{\mathbf{N}}$ which are not θ-u.d. has capacity 0.

Proof (a) For $k \geq 1$, let $f : [0,1]^k \to \mathbb{R}$ be in C^1 and put

$$g = f(X_0, \ldots, X_{k-1})$$

If $N = pk + q$, where $q \in \{0, \ldots, k-1\}$,

$$\frac{1}{N}\sum_{n=0}^{N-1}(g \circ \theta^n - \mathbb{E}(g)) = \frac{1}{N}\left[\sum_{h=0}^{k-1}\sum_{m=0}^{p-1}(g - \mathbb{E}(g)) \circ \theta^{mk+h} + \sum_{i=pk}^{N-1}(g - \mathbb{E}(g)) \circ \theta^i\right].$$

The first term is

$$\frac{p}{N}\sum_{h=0}^{k-1}\frac{1}{p}\sum_{m=0}^{p-1}(g - \mathbb{E}(g)) \circ \theta^{mk+h},$$

and from the previous proposition it follows that this term converges to zero q.e. since, for $h = 0, \ldots, k-1$, the sequence

$$((g - \mathbb{E}(g)) \circ \theta^{mk+h})_{m \geq 0}$$

is orthogonal in $\mathbb{D}$. The remaining term converges to zero everywhere because g is bounded.

(b) Now let k run through the natural numbers. There exists a countable set D of functions g like that in (a), which is convergence-determining in $[0,1]^{\mathbf{N}}$. Since, for any g in D,

$$\frac{1}{N}\sum_{n=0}^{N-1} g \circ \theta^n \to \mathbb{E}(g) \qquad \text{q.e.}$$

it follows that

$$\frac{1}{N}\sum_{n=0}^{N-1}\delta_{\theta^n(\omega)} \to \lambda_1^{\otimes \mathbf{N}}$$

outside a c-null set. □

Comment

Let G be the set of points $\omega \in [0,1]^{\mathbf{N}}$ that are θ-u.d. for $\lambda_1^{\otimes \mathbf{N}}$, that is,

$$\lim_{N\uparrow\infty} \frac{1}{N} \sum_{n=0}^{N-1} \delta_{\theta^n(\omega)} = \lambda_1^{\otimes \mathbf{N}} \qquad \text{narrowly.}$$

The points in G are also called generic points, for the shift according to the measure $\lambda_1^{\otimes \mathbf{N}}$. G has its complement of zero capacity for any capacity satisfying the hypotheses of the preceding proposition. This means intuitively that this set is strictly bigger than the set of points $\omega = (\omega_0, \omega_1, \ldots)$ whose coordinates $\omega_0, \omega_1, \ldots$ are "randomly and independently" drawn.

To see this more explicitly, let us consider the space of Brownian motion equipped with the capacity associated with the Ornstein–Uhlenbeck semi-group. We represent the Brownian motion by a series with respect to an orthonormal basis χ_n of $L^2(\mathbb{R}_+, dx)$ by

$$B(t) = \sum_{n=0}^{\infty} g_n \psi_n(t) \qquad \text{with} \qquad \psi_n(t) = \int_0^t \chi_n(s)\, ds,$$

where the g_n are d-dimensional Gaussian independent reduced random variables (cf. Chapter 5, Section F and Bouleau and Hirsch [1991]). The Brownian motion is represented on $(\mathbb{R}, \mathscr{B}(\mathbb{R}), N(0,1))^{\otimes \mathbf{N}}$ and therefore also on $([0,1], \mathscr{B}[0,1], (\lambda_1))^{\otimes \mathbf{N}}$. The capacity c associated with the Ornstein–Uhlenbeck semi-group is of the type found in Proposition A.8.9 and by a result of Konô [1984], for $d \leq 4$,

$$c\{\omega : \varliminf_{t\to\infty} |B_t(\omega)| < 1\} > 0.$$

If a sample path is "randomly" drawn it is well known that it is transient for $d \geq 3$. Thus a point in G that corresponds to a recurrent path for the Brownian motion of dimension 3 or 4 is not "random."

Explicit sequences in G are known [Rauzy, 1981], which are given by an algorithm (and are therefore not random). But no such "deterministic" sequence at present is known which is significantly faster than random sequences and could play the same role for the shift as low-discrepancy sequences play for finite-dimensional integration.

B REPRESENTATIONS OF STATIONARY FIELDS

This section is an introduction to the simulation of stationary processes and stationary fields. For more details we refer to Azencott and Dacunha-

Castelle [1984], Shiryayev [1984], Rosenblatt [1985], Duflo [1990], and Preumont [1990].

Generally, methods for the statistical identification of stationary processes or fields are dealt with in a wealth of literature, which provides representation and convergence theorems that are very useful for simulation. We shall not tackle such methods here.

Throughout this section, except in Section B.7, random processes and fields will be centered (zero mean) and have finite second-order moments. Thus we confine ourselves to the L^2-theory. Extensions to more general situations are often possible but we do not discuss them here. We shall deal with complex-valued square integrable functions or random variables, unless stated otherwise.

B.1 Wiener Integral

Let $(F, \mathscr{F})$ be a measurable space and $\mathscr{F}_0$ an algebra of subsets that generates $\mathscr{F}$.

Definition B.1.1 A function Z from $\mathscr{F}_0$ to $L^2(\Omega, \mathscr{A}, \mathbb{P})$ is a (centered) orthogonal stochastic measure (OSM) if

(i) for every A in $\mathscr{F}_0$, $\mathbb{E}Z(A) = 0$;

(ii) for every pair A and B of disjoint sets in $\mathscr{F}_0$,

$$\mathbb{E}[Z(A)\overline{Z}(B)] = 0 \qquad \text{and} \qquad Z(A \cup B) = Z(A) + Z(B);$$

(iii) for every sequence (A_n) in $\mathscr{F}_0$ with $A_n \downarrow \varnothing$,

$$\mathbb{E}|Z(A_n)|^2 \to 0, \qquad n \to \infty.$$

Put $\mu_0(A) = \mathbb{E}|Z(A)|^2$ for $A \in \mathscr{F}_0$. The set function μ_0 is a finite measure and consequently, by Caratheodory's theorem, can be extended to $(F, \mathscr{F})$. We denote the resulting measure by μ.

Let $\mathscr{S}$ be the algebra of complex-valued elementary $\mathscr{F}_0$-measurable functions $f = \sum_{j=1}^{n} \alpha_j 1_{A_j}$, where the A_j are disjoint sets in $\mathscr{F}_0$. We define the random variable

$$\int f(\lambda) Z(d\lambda) = \sum_{j=1}^{n} \alpha_j Z(A_j).$$

B.1.2 Exercise
Show that $\mathscr{L}$ is a dense subset of $L^2(F, \mathscr{F}, \mu)$. [*Hint*: The family $\mathscr{C} = \{A : 1_A \in \overline{\mathscr{L}}\}$ is a monotone class containing $\mathscr{F}_0$.]

Lemma B.1.3 The mapping $J: f \rightsquigarrow \int f(\lambda)Z(d\lambda)$ is a linear isometry from $\mathscr{L}$, endowed with the scalar product induced by $L^2(F, \mathscr{F}, \mu)$, into $L^2(\Omega, \mathscr{A}, \mathbb{P})$.

Proof Let f, g be in $\mathscr{L}, f = \sum_{i=1}^{n} \alpha_i 1_{A_i}, g = \sum_{j=1}^{m} \beta_j 1_{B_j}$.

$$\mathbb{E}\left[\int f(\lambda)Z(d\lambda)\overline{\int g(\lambda Z(d\lambda)}\right] = \sum_{i,j} \alpha_i \beta_j \mu(A_i \cap B_j)$$

$$= \int f(\lambda)\overline{g(\lambda)}\mu(d\lambda). \qquad \square$$

Definition B.1.4 The continuous linear map J from a dense subset $\mathscr{L}$ of $L^2(F, \mathscr{F}, \mu)$ to $L^2(\Omega, \mathscr{A}, P)$, admits a unique extension $\tilde{J}$ from $L^2(F, \mathscr{F}, \mu)$ to $L^2(\Omega, \mathscr{A}, P)$, which is a one-to-one homomorphism of Hilbert spaces: we call it the Wiener integral or the stochastic integral, and still write $\int f(\lambda)Z(d\lambda)$.

So, for f, g in $L^2(F, \mathscr{F}, \mu)$, we have

- $\int f(\lambda)Z(d\lambda) \in L^2(\Omega, \mathscr{A}, \mathbb{P})$
- $\mathbb{E}\int f(\lambda)Z(d\lambda) = 0$
- $\mathbb{E}[\int f(\lambda)Z(d\lambda)\overline{\int g(\lambda)Z(d\lambda)}] = \int f\bar{g}\, d\mu$.

B.1.5 Remark
The mapping $A \rightsquigarrow \int 1_A(\lambda)Z(d\lambda)$ from $\mathscr{F}$ to $L^2(\Omega, \mathscr{A}, P)$ defines an OSM that extends Z and is also denoted by Z. The measure μ is said to be associated with Z.

B.1.6 Remark
From the condition $Z(F) \in L^2(\Omega, \mathscr{A}, P)$ it follows that μ is a finite measure with total mass $\mathbb{E}|Z(F)|^2 < \infty$. Let $(F, \mathscr{F})$ be a measurable space and suppose that there are an increasing sequence $(C_k)_{k \in \mathbf{N}}$ of $\mathscr{F}$-measurable sets with $F = \cup_k C_k$ and, for any $k \in \mathbf{N}^*$, an OSM Z_k with associated measure μ_k on $D_k = C_k \backslash C_{k-1} (D_1 = C_1)$ endowed with the restriction of $\mathscr{F}$. Suppose the Z_k are mutually orthogonal. Sticking together the finite measures μ_k, we obtain a σ-finite measure μ on $(F, \mathscr{F})$. For any f in $L^2(F, \mathscr{F}, \mu)$, put

$$\int f(\lambda)Z(d\lambda) = \sum_k \int f_k(\lambda)Z_k(d\lambda),$$

where f_k is the restriction of f to D_k. The mapping

$$f \rightsquigarrow \int f(\lambda)Z(d\lambda)$$

is still a linear isometry from $L^2(F, \mathscr{F}, \mu)$ into $L^2(\Omega, \mathscr{A}, \mathbb{P})$. In particular if A in $\mathscr{F}$ satisfies $\mu(A) = \sum_k \mu_k(A \cap D_k) < \infty$, then the map Z defined by

$$Z(A) = \sum_k Z_k(A \cap D_k)$$

satisfies the previous additivity and orthogonality properties. We call Z a σ-finite OSM.

B.1.7 Example
Let $B^{(1)}$ and $B^{(2)}$ be independent standard one-dimensional Brownian motions. For $t \in \mathbb{R}$ and $k \geq 1$, put

$$\begin{aligned} B_t &= B_t^{(1)}\mathbf{1}_{\{t\geq 0\}} + B_{-t}^{(2)}\mathbf{1}_{\{t<0\}}, \\ C_k &=]-k, +k], \\ D_k &= C_k \backslash C_{k-1} \qquad (D_1 = C_1). \end{aligned}$$

For $n \geq 1$ and $-\infty < a_1 \leq b_1 \leq \cdots \leq a_n \leq b_n < \infty$, put

$$Z\left(\bigcup_{i=1}^{n}]a_i, b_i]\right) = \sum_{i=1}^{n} (B_{b_i} - B_{a_i}).$$

The restrictions Z_k of Z to D_k define mutually orthogonal OSM, and Z can be extended to a σ-finite OSM with associated measure $\mu(dt) = dt$. For f in $L^2(\mathbb{R}, dt)$, we write $\int f(\lambda)dB_\lambda$ in place of $\int f(\lambda)Z(d\lambda)$.

This construction of a σ-finite OSM can be performed for more general processes than Brownian motions.

Definition B.1.8 Let T be either $\mathbb{R}_+$ or $\mathbb{R}$. A set of (complex-valued) random variables $(Y_\lambda)_{\lambda \in T}$ defined on $(\Omega, \mathscr{A}, \mathbb{P})$ is a *process with orthog-*

onal increments (POI) if

(i) $\mathbb{E}|Y_\lambda|^2 < \infty, \quad \lambda \in T;$
(ii) $\mathbb{E}Y_\lambda = 0, \quad \lambda \in T;$
(iii) $\mathbb{P}$-a.s., the paths $\lambda \rightsquigarrow Y_\lambda(\omega)$ are right-continuous functions on T;
(iv) for every $\lambda_1 \le \lambda_2 \le \lambda_3 \le \lambda_4$ in T,

$$\mathbb{E}[(Y_{\lambda_4} - Y_{\lambda_3})\overline{(Y_{\lambda_2} - Y_{\lambda_1})}] = 0.$$

We deduce easily the L^2-right-continuity from (i), (iii), and (iv). Let

$$\begin{aligned} F(\lambda) &= \mathbb{E}|Y_\lambda - Y_0|^2 \qquad \text{if} \quad \lambda \ge 0 \\ &= -\mathbb{E}|Y_\lambda - Y_0|^2 \qquad \text{if} \quad \lambda < 0 \quad \text{and} \quad T = \mathbb{R}. \end{aligned}$$

The right-continuous nondecreasing function F on T defines a σ-finite measure dF. As above, we put

$$Z\left(\bigcup_{i=1}^{n}]a_i, b_i]\right) = \sum_{i=1}^{n} (Y_{b_i} - Y_{a_i})$$

for $a_1 \le b_1 \le \cdots \le a_n \le b_n$ in T. Then Z defines a σ-finite OSM with associated σ-finite measure dF, and we write $\int f(\lambda) dY_\lambda$ in place of $\int f(\lambda) Z(d\lambda)$ for any f in $L^2(T, dF)$.

B.1.9 Example

Let Z be a σ-finite OSM in $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ whose associated measure μ is a Radon measure. For x in $\mathbb{R}^d$, put

$$Y_x = Z\left(\prod_{i=1}^{d}]0 \wedge x_i, 0 \vee x_i]\right) \prod_{i=1}^{d} \operatorname{sign}(x_i).$$

Then $(Y_x)_{x \in \mathbb{R}^d}$ is said to be an *orthogonal random field* (ORF). The multidimensional increments of Y_x corresponding to disjoint d-dimensional intervals are othogonal random variables.

B.1.10 Example

The notion of OSM is often more accessible than that of ORF. Let (φ_n) be an orthonormal basis in $L^2(\mathbb{R}^d, dx)$ and (G_n) be a centered orthonormal sequence in $L^2(\Omega, \mathscr{A}, \mathbb{P})$. For f in $L^2(\mathbb{R}^d, dx)$, put

$$J(f) = \sum_n < f, \varphi_n > G_n.$$

Then J is a linear isometry, and $Z(A) = J(\mathbf{1}_A)$ defines an OSM in $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. If the φ_n are real-valued and the G_n are independent standard normal variables, then $J(f)$ is real Gaussian for any real-valued function f; in fact, the associated ORF has independent increments (*Brownian sheet*).

To handle complex Gaussian processes we must make clear what we mean by a complex Gaussian variable.

Definition B.1.11 A complex random variable $G = G_1 + iG_2$ is complex Gaussian if G_1 and G_2 are independent (centered) real Gaussian variables with same variance σ^2. A vector $X = (X_1, \ldots, X_k)$ is *complex Gaussian* if, for every $\lambda_1, \ldots, \lambda_k$ in $\mathbf{C}$, the random variable $\lambda_1 X_1 + \cdots + \lambda_k X_k$ is complex Gaussian.

In Example B.1.10, if the random variables G_n are independent complex Gaussian with $\sigma^2 = 1$ and (φ_n) is an orthonormal basis of the complex space $L^2(\Omega, \mathscr{A}, \mathbb{P})$, then $J(f)$ is complex Gaussian for every f in the complex space $L^2(\mathbb{R}^d, dx)$.

B.2 Spectral Representation of Stationary (Wide Sense) Fields

Definition B.2.1 A (centered) *random field* X is a measurable map from $\mathbb{R}^d$ to $L^2(\Omega, \mathscr{A}, \mathbb{P})$. It is called *stationary in the wide sense* if for x, y in $\mathbb{R}^d$,

$$\mathbb{E}[X_{x+y}\overline{X}_y] = \mathrm{Cov}\,(X_{x+y}, X_y)$$

does not depend on y. We shall write K_X for the function

$$K_X(x) = \mathrm{Cov}\,(X_{x+y}, X_y)$$

which we call the *covariance function* of X.

B.2.2 Example

Let $\lambda_1, \ldots, \lambda_n$ be n distinct points in $\mathbb{R}^d$ and $\xi_1, \ldots, \xi_n$ pairwise orthogonal (centered) random variables in $L^2(\Omega, \mathscr{A}, \mathbb{P})$. Put

$$X_x = \sum_{j=1}^{n} e^{i\langle \lambda_j, x\rangle} \xi_j, \qquad x \in \mathbb{R}^d.$$

Then X is stationary in the wide sense with covariance function

$$K_X(x) = \sum_{j=1}^{n} e^{i\langle \lambda_j, x\rangle} \|\xi_j\|_2^2.$$

We define an OSM on $\mathscr{B}(\mathbb{R}^d)$

$$Z(A) = \sum_{j=1}^{n} \xi_j \delta_{\lambda_j}(A)$$

with associated measure $\mu = \sum_{j=1}^{n} \|\xi_j\|_2^2 \delta_{\lambda_j}$; we check the relations

$$X_x = \int e^{i\langle \lambda, x\rangle} Z(d\lambda)$$

$$K_X(x) = \int e^{i\langle \lambda, x\rangle} \mu(d\lambda).$$

This construction can easily be extended to the infinite case

$$X_x = \sum_{j=1}^{\infty} e^{i\langle \lambda_j, x\rangle} \xi_j$$

provided $\sum_{j=1}^{\infty} \|\xi_j\|_2^2 < \infty$.

More generally, the following statement is easily checked.

Proposition B.2.3 Let Z be an orthogonal stochastic measure defined on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. Then the random field

$$X_x = \int e^{i\langle \lambda, x\rangle} Z(d\lambda)$$

is stationary in the wide sense and L^2-continuous.

The converse property is a striking fact.

Theorem B.2.4 Let $(X_x)_{x\in\mathbb{R}^d}$ be a stationary random field that is L^2-continuous in the wide sense. There exists a unique OSM Z_X on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ with associated finite measure μ_X such that

$$X_x = \int e^{i\langle\lambda,x\rangle} Z_X(d\lambda)$$

$$K_X(x) = \int e^{i\langle\lambda,x\rangle} \mu_X(d\lambda)$$

for every x in $\mathbb{R}^d$.

Proof The covariance function is a nonnegative definite function:

$$\sum_{i,j=1}^{n} K_X(x_i - x_j)\alpha_i\overline{\alpha}_j = \left\| \sum \alpha_i X_{x_i} \right\|_2^2 \geq 0,$$

for all finite sequences $x_1, \ldots, x_n$ in $\mathbb{R}^d$ and $\alpha_1, \ldots, \alpha_n$ in $\mathbf{C}$; it is also continuous. It follows by Bochner's theorem that it is the Fourier transform of a finite measure μ_X in $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. The mapping

$$\sum_{j=1}^{n} \alpha_j X_{x_j} \overset{\varphi}{\rightsquigarrow} \sum_{j=1}^{n} \alpha_j e^{i\langle\lambda,x_j\rangle}$$

is a well-defined isometry from a linear subspace of $L^2(\Omega, \mathscr{A}, \mathbb{P})$ onto a linear subspace of $L^2(\mathbb{R}^d, \mu_X)$. It can be uniquely extended to the generated closed subspaces. On the left-hand side, this subspace is denoted by L_X^2; on the right-hand side it is the whole space $L^2(\mathbb{R}^d, \mu_X)$. If, for $f \in L^2(\mathbb{R}^d, \mu_X)$, we put

$$\int f(\lambda) Z_X(d\lambda) = \varphi^{-1}(f),$$

then we have

$$\mathbb{E}[|\varphi^{-1}(f)|^2] = \|f\|^2_{L^2(\mu_X)} = \int |f|^2 \mu_X(d\lambda)$$

and, for $e_x(\lambda) = e^{i\langle\lambda,x\rangle}$,

$$\varphi^{-1}(e_x) = X_x.$$

It follows that Z_X is an OSM with associated measure μ_X, and that it is the unique solution. □

Definition B.2.5 The measure μ_X is called the *spectral measure* of X. Its density, if it has one, is called the *spectral density* of X.

B.3 Linear Transformations of Stationary Fields

Given a field $X = (X_x)_{x\in\mathbb{R}^d}$ that is stationary in the wide sense, we can get other stationary fields $Y = (Y_y)_{y\in\mathbb{R}^d}$ by various linear operations (derivations, convolutions, etc.), which commute with translations in the parameter space. These operations are not, however, defined for all stationary fields unless we use the theory of distributions and generalized functions [I.M. Gelfand, N.Y. Vilenkin]. We shall confine ourselves to the L^2-theory.

Definition B.3.1 Let X be an L^2-continuous stationary random field with spectral representation

$$X_x = \int e^{i\langle\lambda,x\rangle} Z_X(d\lambda)$$

and covariance function.

$$K_X(x) = \int e^{i\langle\lambda,x\rangle} \mu_X(d\lambda).$$

A linear transformation T defined on a vector space of stationary random fields that includes X is said to be *X-admissible* if there exists $C \in L^2(\mathbb{R}^d, \mu_X)$ such that

$$(*) \qquad T(X)_x = \int e^{i\langle\lambda,x\rangle} C(\lambda) Z_X(d\lambda), \qquad x \in \mathbb{R}^d.$$

The function C is called the *transfer function*, or the *gain*, or the *spectral characteristic* of the operation.

We easily get the following result.

Proposition B.3.2 The random field $Y = T(X)$ defined by (*) is (wide sense) stationary with covariance function

$$K_Y(x) = \int e^{i\langle \lambda, x\rangle} |C(\lambda)|^2 \mu_X(d\lambda)$$

and the random fields X and Y are stationarily correlated:

$$\mathbb{E}[Y_{x+y}\overline{X}_y] = \int e^{i\langle \lambda, x\rangle} C(\lambda)\mu_X(d\lambda).$$

This last function of x is called the *cross covariance function* of Y and X and is denoted by K_{YX}.

B.3.3 Example

(Band-pass filter.) For $d = 1$, take $C(\lambda) = 1_{[\lambda_1,\lambda_2]}(\lambda)$ for $-\infty < \lambda_1 < \lambda_2 < +\infty$. The previous proposition shows that the spectral measure of Y is equal to the restriction of the spectral measure of X to the interval $[\lambda_1, \lambda_2]$: so we have a band-pass filter that keeps only the waveband $[\lambda_1, \lambda_2]$.

B.3.4 Example

(Linear combinations of parameter translations.) Let

$$Y_x = \sum_{j=1}^{k} \alpha_j X_{x+x_j}, \qquad x \in \mathbb{R}^d$$

for $\alpha_1, \ldots, \alpha_k$ in $\mathbf{C}$ and $x_1, \ldots, x_k$ in $\mathbb{R}^d$. The transfer function of the transformation taking X to Y is

$$C(\lambda) = \sum_{j=1}^{k} \alpha_j e^{i\langle \lambda, x_j\rangle}.$$

B.3.5 Example

(Moving averages.) Let X be an L^2-continuous (wide sense) stationary random field with associated OSM Z_X, and let ν be a finite measure in $\mathbb{R}^d$. Put

$$C(\lambda) = \int e^{-i\langle \lambda, x\rangle} \nu(dx), \qquad \lambda \in \mathbb{R}^d.$$

From Fubini's theorem for Wiener integrals (see Exercise B.3.6), it follows that

$$\int \left(\int e^{i\langle \lambda, x-y\rangle} Z_X(d\lambda) \right) d\nu(y) = \int e^{i\langle \lambda, x\rangle} \left(\int e^{-i\langle \lambda, y\rangle} \nu(dy) \right) Z_X(d\lambda).$$

The bounded function C is the transfer function of the operation that takes X to the stationary random field

$$Y_x = \int X_{x-y} d\nu(y).$$

This integral is a Bochner integral in the Hilbert space $L^2(\Omega, \mathscr{A}, \mathbb{P})$.

B.3.6 Exercise

(Fubini's theorem.) Let Z be an OSM with associated finite measure μ. Let ν be a finite measure in $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ and $\varphi \in L^2(\mathbb{R}^d \times \mathbb{R}^d, \mu \otimes \nu)$. Show that

$$\int \left(\int \varphi(\lambda, x) Z(d\lambda) \right) \nu(dx) = \int \left(\int \varphi(\lambda, x) \nu(dx) \right) Z(d\lambda).$$

[*Hint*: First show this relation holds for $\varphi(\lambda, x) = \theta(\lambda)\psi(x)$. Then approximate φ in $L^2(\mu \otimes \nu)$ by finite sums

$$\varphi_n(\lambda, x) = \sum_k \theta_k(\lambda)\psi_k(x).$$

Show that

$$\int \varphi_n(\lambda, x) \nu(dx) \to \int \varphi(\lambda, x) \nu(dx) \qquad \text{in} \quad L^2(\mathbb{R}^d, \mu)$$

and deduce that in $L^2(\Omega, \mathscr{A}, \mathbb{P})$,

$$\int \left(\int \varphi_n(\lambda, x) \nu(dx) \right) Z(d\lambda) \to \int \left(\int \varphi(\lambda, x) \nu(dx) \right) Z(d\lambda).$$

Then show that in $L^2(\Omega, \mathscr{A}, \mathbb{P})$,

$$\int \left(\int \varphi_n(\lambda, x) Z(d\lambda) \right) \nu(dx) \to \int \left(\int \varphi(\lambda, x) Z(d\lambda) \right) \nu(dx).]$$

B.3.7 Example

(Differentiation.) Let X be an L^2-continuous stationary random field such that its spectral measure μ_X satisfies the condition

$$\int |\lambda|^2 \mu_X(d\lambda) < \infty.$$

For $k = 1, \ldots, d$, the X-admissible transformation with transfer function

$$C_k(\lambda) = i\lambda_k$$

maps the stationary process X to its derivative process $(\partial X_x/\partial x_k)_{x \in \mathbb{R}^d}$. If, in addition,

$$\int |\lambda|^4 \mu_X(d\lambda) < \infty,$$

then X is continously twice differentiable with respect to the L^2-norm. For $b = (b_i)_{i=1,\ldots,d}$, $c = (c_{kl})_{k,l=1,\ldots,d}$, if we put

$$Y_x = aX_x + \sum_{j=1}^{d} b_j \frac{\partial X_x}{\partial x_j} + \sum_{k,l=1}^{d} c_{kl} \frac{\partial^2 X_x}{\partial x_k \partial x_l},$$

then Y has the spectral representation

$$Y_x = \int e^{i\langle \lambda, x \rangle} (a + i\langle \lambda, b \rangle - \langle \lambda, c\lambda \rangle) Z_X(d\lambda).$$

B.4 Second-Order Linear Differential Equations

Suppose first that $d = 1$.

Proposition B.4.1 Let X be an L^2-continuous stationary process on $\mathbb{R}$ and let a_1, a_2 be complex numbers that are not purely imaginary. Then the equation

$$(*) \qquad Y_t'' - (a_1 + a_2) Y_t' + a_1 a_2 Y_t = X_t$$

has a unique continuously twice L^2-differentiable stationary solution Y.

Proof Suppose that the process Y is a solution with spectral representation

$$Y_t = \int e^{i\lambda t} Z_Y(d\lambda), \qquad t \in \mathbb{R}.$$

It follows from the hypotheses that

$$\int \lambda^4 \mu_Y(d\lambda) < \infty$$

and

$$Y_t' = \int i\lambda e^{i\lambda t} Z_Y(d\lambda)$$

$$Y_t'' = \int (-\lambda^2) e^{i\lambda t} Z_Y(d\lambda).$$

Thus we have

$$X_t = \int e^{i\lambda t}(a_1 a_2 - (i\lambda)(a_1 + a_2) - \lambda^2) Z_Y(d\lambda);$$

from the uniqueness of the spectral representation, it follows that

$$Z_X(d\lambda) = (a_1 a_2 - (i\lambda)(a_1 + a_2) - \lambda^2) Z_Y(d\lambda).$$

Put

$$C(\lambda) = \frac{1}{(i\lambda)^2 - (i\lambda)(a_1 + a_2) + a_1 a_2}$$

$$= \begin{cases} \dfrac{1}{a_1 - a_2}\left[\dfrac{1}{i\lambda - a_1} - \dfrac{1}{i\lambda - a_2}\right] & \text{if } a_1 \neq a_2 \\ \dfrac{1}{(i\lambda - a_1)^2} & \text{if } a_1 = a_2. \end{cases}$$

Then $Z_Y(d\lambda) = C(\lambda) Z_X(d\lambda)$. Conversely, as $C \in L^2(\mathbb{R}, \mu_X)$, this relation defines an OSM Z_Y and it is easy to check that

$$Y_t = \int e^{i\lambda t} Z_Y(d\lambda)$$

is the solution to the problem. □

To see what can happen if a_1 or a_2 is purely imaginary, let us consider the equation

$$(**) \qquad Y_t' - i\alpha Y_t = X_t \qquad t \in \mathbb{R}$$

where $\alpha \in \mathbb{R}$ and Y is continuously L^2-differentiable. If Y is stationary, then

$$(\lambda - \alpha)^2 \mu_Y(d\lambda) = \mu_X(d\lambda),$$

so we must have

$$(***) \qquad \int \frac{1}{(\lambda - \alpha)^2} \mu_X(d\lambda) < \infty.$$

If this condition is fulfilled, the general stationary solution of (**) is given by

$$Y_t = \xi e^{i\alpha t} + \int \frac{e^{i\lambda t}}{i(\lambda - \alpha)} Z_X(d\lambda),$$

where ξ is any (centered) random variable in $L^2(\Omega, \mathscr{A}, \mathbb{P})$ which is orthogonal to the subspace L_X^2. If, more generally, we do not require the condition (***) to be satisfied, then any process

$$\begin{aligned} Y_t &= Y_0 e^{i\alpha t} + e^{i\alpha t} \int_0^t e^{-i\alpha s} X_s ds \\ &= Y_0 e^{i\alpha t} + \int \frac{e^{i\lambda t} - e^{i\alpha t}}{i(\lambda - \alpha)} Z_X(d\lambda), \end{aligned}$$

where $Y_0 \in L^2(\Omega, \mathscr{A}, P)$, is a solution of (**), although not in general a stationary solution.

B.4.2 Example

The response of a mechanical system with one degree of freedom to the force X_t is governed by the equation

$$Y_t'' + 2\xi\omega_0 Y_t' + \omega_0^2 Y_t = X_t,$$

where $\omega_0 > 0$ and $0 < \xi < 1$. The transfer function C taking X to Y is

$$C(\lambda) = \frac{1}{(i\lambda)^2 + 2\xi\omega_0 i\lambda + \omega_0^2}.$$

Notice that the formula

$$Y_t = \int e^{i\lambda t} C(\lambda) Z_X(d\lambda)$$

defines a continuously L^2-differentiable stationary process, not only in case of finite spectral measure μ_X but also when μ_X is Lebesgue measure; in this case the generalized process X is called *white noise* and Z_X is a σ-finite OSM. The spectral measure of Y is

$$\mu_Y(d\lambda) = \frac{d\lambda}{(\omega_0^2 - \lambda^2)^2 + 4\xi^2\omega_0^2\lambda^2}$$

and its covariance function

$$K_Y(t) = \frac{\pi e^{-\omega_0\xi|t|}}{2\omega_0^3}\left[\frac{1}{\xi}\cos\omega_0\sqrt{1-\xi^2}|t| + \frac{1}{\sqrt{1-\xi^2}}\sin\omega_0\sqrt{1-\xi^2}|t|\right].$$

We handle partial differential equations for stationary random fields in a similar way. For instance, the equation

$$aY + i\langle b, \nabla Y\rangle - \langle \nabla Y, c\nabla Y\rangle = X$$

has a continuously twice L^2-differentiable stationary solution Y for any L^2-continuous stationary field X, provided that the matrix c is symmetric positive definite and the function

$$a + i\langle b, \lambda\rangle - \langle \lambda, c\lambda\rangle$$

is non-zero over the whole space $\mathbb{R}^d$. Otherwise, we need additional hypotheses on the spectral measure of X for the equation to have a stationary solution.

B.5 Multidimensional Processes and Fields

In practice, it is often necessary to make use of multidimensional processes and fields. The theory is very similar, and we shall only give the notation.

Definition B.5.1 An *n-dimensional (wide sense) stationary field* $X =$

$(X_x)_{x\in\mathbb{R}^d}$ is a collection of n stationary, pairwise stationarily connected fields, say, $X_x^1,\dots,X_x^n$.

This means that the matrix $\mathbb{E}[X_{x+y}^t\overline{X}_y]$ does not depend on y in $\mathbb{R}^d$. It is called the *covariance function* K_X of the multidimensional field X.

Definition B.5.2 A so-called *vector-valued (centered) orthogonal stochastic measure* (VOSM) Z on a measurable space $(F,\mathscr{F})$ is a mapping from $\mathscr{F}$ to $(L^2(\Omega,\mathscr{A},P))^n$ such that

(i) for every A in $\mathscr{F}$, $\mathbb{E}Z(A)=0$;

(ii) for every pair of disjoint sets A and B in $\mathscr{F}$,

$$\mathbb{E}[Z(A)^t\overline{Z(B)}]=0 \qquad \text{and} \qquad Z(A\cup B)=Z(A)+Z(B);$$

(iii) for every sequence (A_n) in $\mathscr{F}$ with $A_n\downarrow\varnothing$,

$$\mathbb{E}\|Z(A_n)\|_2^2\to 0, \qquad n\to\infty.$$

A matrix measure

$$\mu(A)=\mathbb{E}[Z(A)^t\overline{Z(A)}], \qquad A\in\mathscr{F},$$

is associated with the VOSM Z. For any A in $\mathscr{F}$, the matrix $(\mu_{ij}(A))_{i,j=1,\dots,n}$ is Hermitian and nonnegative definite.

The trace $\tau=\sum_{j=1}^d\mu_{jj}$ of μ is a nonnegative measure in $\mathbb{R}^d$ and the space $L^2(\mathbb{R}^d,\tau)$ is denoted by $L^2(\mathbb{R}^d,\mu)$.

A stochastic integral may be defined as in Definition B.1.4. If X is a multidimensional L^2 continuous stationary random field, then there exists a unique VOSM Z_X on $(\mathbb{R}^d,\mathscr{B}(\mathbb{R}^d))$ such that

$$X_x=\int e^{i\langle\lambda,x\rangle}Z_X(d\lambda);$$

the matrix measure μ_X associated Z_X is called the spectral measure of X, and

$$K_X(x)=\int e^{i\langle\lambda,x\rangle}\mu_X(d\lambda).$$

B.6 Real-Valued Fields

Proposition B.6.1 A (wide sense) stationary random field X is real-valued iff the associated OSM Z_X satisfies

$$Z_X(-A) = \overline{Z_X(A)}$$

for every $A \in \mathscr{B}(\mathbb{R}^d)$. In this case, $\mu_X(-A) = \mu_X(A)$.

Proof Assume first that $Z_X(-A) = \overline{Z_X(A)}$ for every A in $\mathscr{B}(\mathbb{R}^d)$. Then

$$\mu_X(-A) = \mathbb{E}[Z_X(-A)\overline{Z_X(-A)}] = \mu_X(A).$$

For every elementary $f = \sum_{j=1}^n \alpha_j \mathbf{1}_{A_j}$,

$$\overline{\int f(\lambda) Z_X(d\lambda)} = \overline{\sum_{j=1}^n \alpha_j Z_X(A_j)} = \sum_{j=1}^n \overline{\alpha}_j Z_X(-A_j) = \int \overline{f(-\lambda)} Z_X(d\lambda)$$

and by density this equality holds for every f in $L^2(\mathbb{R}^d, \mu_X)$. Thus we have

$$\overline{X_x} = \overline{\int e^{i\langle \lambda, x\rangle} Z_X(d\lambda)} = \int e^{i\langle \lambda, x\rangle} Z_X(d\lambda) = X_x.$$

Conversely, suppose that X is real-valued. Then for every x in $\mathbb{R}^d$,

$$K_X(-x) = K_X(x)$$

and it follows from the uniqueness of the spectral measure that, for every A in $.\mathscr{B}(\mathbb{R}^d)$,

$$\mu_X(A) = \mu_X(-A).$$

If $f(\lambda) = \sum_{j=1}^n \alpha_j e^{i\langle \lambda, x_j\rangle}$, then

$$\overline{\int f(\lambda) Z_X(d\lambda)} = \overline{\sum_{j=1}^n \alpha_j X_{x_j}} = \sum_{j=1}^n \overline{\alpha}_j X_{x_j} = \int \overline{f(-\lambda)} Z_X(d\lambda),$$

and by taking limits this equality extends to every f in $L^2(\mathbb{R}^d, \mu_X)$. In particular,

$$\overline{Z_X(A)} = Z_X(-A), \qquad A \in \mathscr{B}(\mathbb{R}^d). \qquad \square$$

Let us return to the real Brownian motion B in Example B.1.7. If $f \in L^2(\mathbb{R}, d\lambda)$ is real-valued, the stationary process

$$X_t = \int e^{i\lambda t} f(\lambda) dB_\lambda$$

is not real-valued, its real part is not stationary, and the random variables X_t are not complex Gaussian in the sense of Definition B.1.11. In fact, it is more convenient to deal with a complex-valued Brownian motion.

Definition B.6.2 A complex-valued process W on $\mathbb{R}_+$ is a *complex standard Brownian motion* on $\mathbb{R}_+$ if there exist two independent real standard Brownian motions $B^{(1)}$ and $B^{(2)}$ such that

$$W_\lambda = \frac{1}{\sqrt{2}} (B^{(1)}_\lambda + iB^{(2)}_\lambda), \qquad \lambda \geq 0.$$

Such a process W is a complex-valued POI. For any f in $L^2(\mathbb{R}_+, d\lambda)$,

$$Y_t = \int e^{i\lambda t} f(\lambda) dW_\lambda$$

is a complex stationary process. Finally, if we put

$$\tilde{W}_\lambda = W_\lambda 1_{\{\lambda \geq 0\}} - \overline{W}_{-\lambda} 1_{\{\lambda < 0\}},$$

then it is easily checked that $\tilde{W}$ is a POI on $\mathbb{R}$ and, for any real even f in $L^2(\mathbb{R}, d\lambda)$,

$$\tilde{X}_t = \int e^{i\lambda t} f(\lambda) d\tilde{W}_\lambda$$

is a real stationary process with spectral (density $f(\lambda))^2$.

B.6.3 Exercise
Check that

$$\mathbb{E}e^{i\,\mathrm{Re}(zX_\lambda)} = \exp\left\{-\frac{|z|^2}{4}\lambda\right\},$$

for any z in $\mathbf{C}$ and λ in $\mathbb{R}_+$.

B.6.4 Exercise
(Conformal local martingales.) If H is any analytic function in $\mathbf{C}$, show that $U_\lambda = \mathrm{Re}[H(W_\lambda)]$ and $V_\lambda = \mathrm{Im}[H(W_\lambda)]$ are local martingales such that $\langle U\rangle_\lambda = \langle V\rangle_\lambda$ and $\langle U, V\rangle_\lambda = 0$. [*Hint*: Combine Ito's formula with the Cauchy–Riemann equations to find that $dH(W_\lambda) = H'(W_\lambda)dW_\lambda$.]

B.7 Strictly Stationary Fields

Definition B.7.1 A (centered) random field $X = (X_x)_{x\in\mathbb{R}^d}$ is *stationary in the strict sense* if its finite-dimensional distributions are invariant under all shifts of parameter; that is, $(X_{x_1+y}, \ldots, X_{x_n+y})$ has the same distribution as $(X_{x_1}, \ldots, X_{x_n})$ for every $n \geq 1$ and $x_1, \ldots, x_n, y$ in $\mathbb{R}^d$.

Definition B.7.2 A real-valued random field $(X_x)_{x\in\mathbb{R}^d}$ is a *real Gaussian field* if for all $x_1, \ldots, x_n$ in $\mathbb{R}^d$, any real linear combination of $X_{x_1}, \ldots, X_{x_n}$ is a real Gaussian variable.

Recall that we consider only centered random fields and variables, so the finite-dimensional distributions of a real Gaussian field are completely determined by the correlation function $\mathbb{E}[X_{x+y}X_y]$. Thus, if a real Gaussian field is stationary in the wide sense, it is also stationary in the strict sense. For instance, the process $\tilde{X}$ in B.6.2 is real Gaussian stationary.

Definition B.7.3 In the same way, a complex-valued random field $(X_x)_{x\in\mathbb{R}^d}$ is a *complex Gaussian field* if for all $x_1, \ldots, x_n$ in $\mathbb{R}^d$, every complex linear combination of $X_{x_1}, \ldots, X_{x_n}$ is a complex Gaussian variable.

A complex wide sense stationary Gaussian field is also stationary in the strict sense. The process Y in Definition B.6.2 is complex Gaussian stationary.

We obtain more general examples of processes that are stationary in the strict sense by introducing processes with independent increments. They are more general than Brownian motion.

For the moment we shall adopt the following definition, which is convenient for our purpose. A wider definition will be given later in Section D.

Definition B.7.4 Let T be $\mathbb{R}_+$ or $\mathbb{R}$. A collection of random variables $Y = (Y_\lambda)_{\lambda \in T}$ is a (square integrable centered) *process with stationary independent increments* (PSII) if

(i) $\mathbb{E}|Y_\lambda|^2 < \infty, \quad \lambda \in T;$
(ii) $\mathbb{E}Y_\lambda = 0, \quad \lambda \in T;$
(iii) $\mathbb{P}$-a.s., the paths $\lambda \rightsquigarrow Y_\lambda(\omega)$ are right-continuous on T;
(iv) for $\lambda_0 < \cdots < \lambda_n$ in T, the increments $Y_{\lambda_1} - Y_{\lambda_0}, \ldots, Y_{\lambda_n} - Y_{\lambda_{n-1}}$ are mutually independent;
(v) the distribution of $Y_{\lambda+\mu} - Y_\mu$ does not depend on μ, for every λ in T.

It is straightforward to check that a PSII is a POI, and the complex standard Brownian motion in Definition B.6.2 is a PSII. But the POI $\tilde{W}$ is not a PSII on $\mathbb{R}$.

Definition B.7.5 A real-valued process $S = (S_\lambda)_{\lambda \in \mathbb{R}_+}$ is a *subordinator* if

(i) $S_0 = 0$ and a.s. S has nondecreasing right-continuous paths;
(ii) for $\lambda_0 < \cdots < \lambda_n$ in $\mathbb{R}_+$, the increments $S_{\lambda_1} - S_{\lambda_0}, \ldots, S_{\lambda_n} - S_{\lambda_{n-1}}$ are mutually independent;
(iii) for λ in $\mathbb{R}_+$, the distribution of $S_{\lambda+\mu} - S_\mu$ does not depend on μ.

B.7.6 Remark
This terminology is due to the following property, which is called subordination in the Bochner sense (see Section D.7 below for more details). Let $(X_t)_{t \geq 0}$ be a Markov process with state space $(E, \mathscr{E})$ and transition semigroup $(P_t)_{t \geq 0}$. For every bounded $\mathscr{E}$-measurable function f, for $s \leq t$, then

$$\mathbb{E}[f(X_t)|X_u, u \leq s] = P_{t-s}f(X_s) \qquad \mathbb{P}\text{-a.s.}$$

(see Chapter 4). Let S be a subordinator independent of the process X. Then it can be proved that the process $(X_{S_\lambda})_{\lambda \geq 0}$ is still a Markov process, with transition semi-group

$$Q_\lambda = \int_0^\infty P_s \mu_\lambda(ds),$$

where μ_λ is the distribution of S_λ.

B.7.7 Exercise
Verify the assertion of Remark B.7.6.
Since for every $p \geq 0$

$$\mathbb{E}[e^{-pS_{\lambda+\mu}}] = \mathbb{E}[e^{-pS_\lambda}]\mathbb{E}[e^{-pS_\mu}],$$

there exists a continuous nondecreasing function $\varphi : [0, \infty[\to [0, \infty[$ with $\varphi(0) = 0$ such that

$$\mathbb{E}[e^{-pS_\lambda}] = e^{-\lambda\varphi(p)} \qquad p, \lambda \geq 0.$$

The function φ is a *Bernstein function* (see Section D.3 below): for $k \geq 1$, and $p > 0$, we have $(-1)^k(d^k\varphi/dp^k) \leq 0$. In this section, we consider only the case $\varphi'(0) < \infty$, which is equivalent to assuming $\mathbb{E}S_1 < \infty$. From the Lévy–Hincin formula (see Section D.3 below), it follows that there exists $m \geq 0$ and a σ-finite measure μ on $]0, \infty[$ satisfying $\int_0^\infty x\mu(dx) < \infty$ such that

$$\varphi(p) = mp + \int_0^\infty (1 - e^{-px})\mu(dx).$$

Proposition B.7.8 Let $(W_t)_{t\geq 0}$ be a complex standard Brownian motion and $(S_\lambda)_{\lambda\geq 0}$ a subordinator independent of W, with $\varphi'(0) < \infty$. Then $Y_\lambda = W_{S_\lambda}$ is a PSII and, for any $f \in L^2(\mathbb{R}_+, d\lambda)$,

$$X_t = \int e^{i\lambda t} f(\lambda) dY_\lambda, \qquad t \in \mathbb{R}$$

is a strictly stationary process.

Proof (a) The paths of Y are right-continuous because those of S are right-continuous and those of W are continuous. We have

$$\mathbb{E}|Y_\lambda|^2 = \tfrac{1}{2}\mathbb{E}[(B^{(1)}_{S_\lambda})^2 + (B^{(2)}_{S_\lambda})^2] = \mathbb{E}S_\lambda = \lambda\varphi'(0) < \infty.$$

Trivially $\mathbb{E}Y_\lambda = 0$. For $n \geq 1, u_1, \ldots, u_n$ in C and $0 \leq \lambda_0 \leq \lambda_1 \cdots \leq \lambda_n < \infty$, we have

$$\begin{aligned}
&\mathbb{E}\left[\exp\left\{i\,\mathrm{Re}\sum_{j=1}^{n}\overline{u}_j(Y_{\lambda_j}-Y_{\lambda_{j-1}})\right\}\right] \\
&\quad=\mathbb{E}\left[\exp\left\{\frac{i}{\sqrt{2}}\left(\sum_{j=1}^{n}\mathrm{Re}\,u_j(B^{(1)}_{S_{\lambda_j}}-B^{(1)}_{S_{\lambda_{j-1}}})\right.\right.\right. \\
&\qquad\left.\left.\left.+\sum_{j=1}^{n}\mathrm{Im}\,u_j(B^{(2)}_{S_{\lambda_j}}-B^{(2)}_{S_{\lambda_{j-1}}})\right)\right\}\right] \\
&\quad=\mathbb{E}\left[\prod_{j=1}^{n}e^{-(|u_j|^2/4)(S_{\lambda_j}-S_{\lambda_{j-1}})}\right] \\
&\quad=\prod_{j=1}^{n}\exp\left\{-(\lambda_j-\lambda_{j-1})\varphi\left(\frac{|u_j|^2}{4}\right)\right\} \\
&\quad=\prod_{j=1}^{n}\mathbb{E}\left[\exp\left\{i\,\mathrm{Re}\sum_{j=1}^{n}\overline{u}_j(Y_{\lambda_j}-Y_{\lambda_{j-1}})\right\}\right].
\end{aligned}$$

It follows that the process Y has independent increments and the distribution of $Y_{\lambda+\mu}-Y_\mu$ depends only on λ.

(b) For any elementary function $g(\lambda)=\sum_{j=1}^{m}\overline{u}_j\mathbf{1}_{]\lambda_{j-1},\lambda_j]}(\lambda)$, we have seen that

$$\begin{aligned}
\mathbb{E}\left[\exp\left\{i\,\mathrm{Re}\int g(\lambda)dY_\lambda\right\}\right]&=\exp\left\{-\sum_{j=1}^{m}(\lambda_j-\lambda_{j-1})\varphi\left(\frac{|u_j|^2}{4}\right)\right\} \\
&=\exp\left\{-\int_0^\infty\varphi\left(\frac{|g(\lambda)|^2}{4}\right)d\lambda\right\}.
\end{aligned}$$

Note that, for h and k in $L^2(\mathbb{R}_+,d\lambda)$,

$$\mathbb{E}\left|\int h(\lambda)dY_\lambda - \int k(\lambda)dY_\lambda\right|^2 = (\varphi'(0))^2 \int_0^\infty |h-k|^2 d\lambda.$$

For any g in $L^2(\mathbb{R}_+, d\lambda)$, take a sequence of elementary functions (g_n) that converges to g in $L^2(\mathbb{R}_+, d\lambda)$. We find that

$$\begin{aligned}\mathbb{E}\left[\exp\left\{i\,\mathrm{Re}\int g(\lambda)dY_\lambda\right\}\right] &= \lim_n \mathbb{E}[\exp\{i\,\mathrm{Re}\int g_n(\lambda)dY_\lambda\}] \\ &= \lim_n \exp\left\{-\int_0^\infty \varphi\left(\frac{|g_n(\lambda)|^2}{4}\right)d\lambda\right\} \\ &= \exp\left\{-\int_0^\infty \varphi\left(\frac{|g(\lambda)|^2}{4}\right)d\lambda\right\}.\end{aligned}$$

The last convergence is a straightforward consequence of the inequalities $0 \le \varphi'(p) \le \varphi'(0) < \infty$ for $p \ge 0$. For $v_1, \ldots, v_n$ in $\mathbf{C}$ and $t_1, \ldots, t_n$ in $\mathbb{R}$,

$$\begin{aligned}&\mathbb{E}\left[\exp\left\{i\,\mathrm{Re}\sum_{j=1}^n \bar{v}_j X_{t_j}\right\}\right] \\ &\quad= \mathbb{E}\left[\exp\left\{i\,\mathrm{Re}\int\left(\sum_{j=1}^n \bar{v}_j e^{i\lambda t_j}\right)f(\lambda)dY_\lambda\right\}\right] \\ &\quad= \exp\left\{-\int_0^\infty \varphi\left[\left(\left|\sum_{j=1}^n \bar{v}_j e^{i\lambda t_j}\right|^2 \Big/ 4\right)|f(\lambda)|^2\right]d\lambda\right\},\end{aligned}$$

from which it follows that the finite-dimensional distributions of X are invariant under parameter shifts. □

B.7.9 Remark

To get a PSII over $T = \mathbb{R}$ instead of $T = \mathbb{R}_+$, we take two independent copies $Y_\lambda^{(1)}$ and $Y_\lambda^{(2)}$ over $\mathbb{R}_+$ and define

$$Y_\lambda = Y_\lambda^{(1)}\mathbf{1}_{\{\lambda \ge 0\}} - Y_{(-\lambda)-}^{(2)}\mathbf{1}_{\{\lambda<0\}}.$$

An alternative way to obtain strict sense stationary processes is to define moving averages with respect to a PSII. For instance, given a function f in $L^2(\mathbb{R}, d\lambda)$, it is easy, following the proof of Proposition B.7.8, to check that

$$X_t = \int f(t-\lambda) dY_\lambda$$

is a strict sense stationary process. We will draw the same conclusion for another class of PSII.

Let $Y^{(1)}$ and $Y^{(2)}$ be two independent copies of a real-valued PSII on $\mathbb{R}_+$ starting at 0 (i.e., $Y_0^{(1)} = Y_0^{(2)} = 0$). Their exponent function ψ is defined by

$$\mathbb{E}[e^{iuY_\lambda^{(1)}}] = \exp -\lambda\psi(u), \qquad \lambda \geq 0, \quad u \in \mathbb{R};$$

by the Lévy–Hincin formula, there exists $\sigma \geq 0$ and a σ-finite measure ν on $\mathbb{R}^* = \mathbb{R}\backslash\{0\}$ such that

$$\int_{\mathbb{R}^*} b^2 \nu(db) < \infty$$

and

$$\psi(u) = +\tfrac{1}{2}\sigma^2 u^2 - \int_{\mathbb{R}^*} (e^{iub} - 1 - iub)\nu(db).$$

It is well known that every PSII has a version whose paths a.s. have limits from the left, so we may put

$$Y_\lambda = Y_\lambda^{(1)} 1_{\{\lambda \geq 0\}} - Y_{(-\lambda)-}^{(2)} 1_{\{\lambda < 0\}}, \qquad \lambda \in \mathbb{R}.$$

This is a real PSII on the whole line.

Proposition B.7.10 Let f be a real-valued function in $L^2(\mathbb{R}, d\lambda)$. Then the process

$$X_t = \int f(t-\lambda) dY_\lambda$$

is stationary in the strict sense.

Proof If g is a real elementary function with compact support in $\mathbb{R}$,

it is easy to check that

$$E\left[\exp i\int g(\lambda)dY_\lambda\right] = \exp -\int \psi(g(\lambda))d\lambda.$$

As in the proof of Proposition B.7.8, this formula extends to any real function g in $L^2(\mathbb{R}, d\lambda)$. In fact, if $g_n \to g$ in $L^2(\mathbb{R}, d\lambda)$, then

$$\int g_n(\lambda)dY_\lambda \to \int g(\lambda)dY_\lambda \qquad \text{in } L^2(\Omega, \mathscr{A}, \mathbb{P})$$

and

$$\left|\int(\psi(g_n(\lambda)) - \psi(g(\lambda)))d\lambda\right|$$
$$\leq \frac{\sigma^2}{2}\int |g_n^2 - g^2|d\lambda + \int\int_{\mathbb{R}^*} |e^{ibg_n} - e^{ibg} - ib(g_n - g)|d\nu(b)d\lambda.$$

Since there exists $c > 0$ such that

$$|e^{ibg_n} - e^{ibg} - ib(g_n - g)|$$
$$\leq |e^{ib(g_n-g)} - 1 - ib(g_n - g)| + |(e^{ibg} - 1)b(g_n - g)|$$
$$\leq cb^2|g_n - g|^2 + b^2|g||g_n - g|,$$

it follows that

$$\exp -\int \psi(g_n(\lambda))d\lambda \to \exp -\int \psi(g(\lambda))d\lambda$$

and we are done. For $-\infty < t_1 < \cdots < t_n < \infty$ and $u_1, \ldots, u_n$ in $\mathbb{R}$, we find that

$$\mathbb{E}[\exp i(u_1X_{t_1} + \cdots + u_nX_{t_n})] = \mathbb{E}\left[\exp i\int \sum_{j=1}^n u_jf(t_j - \lambda)dY_\lambda\right]$$
$$= \exp -\int \psi\left(\sum_{j=1}^n u_jf(t_j - \lambda)\right)d\lambda.$$

This last expression is clearly invariant under time translation. □

B.8 Space and Time Discretizations

A stationary process may be discretized with respect to frequency or time (space in the case of a random field). We first deal with discretization with respect to time.

(a) Simulating Stationary Processes by Stationary Sequences
We restate briefly the main features of the discrete case.

Definition B.8.1 A complex random sequence $X = (X_n)_{n\in\mathbf{Z}}$ is *stationary in the wide sense* if

(i) $X_n \in L^2(\Omega, \mathscr{A}, \mathbb{P})$ $\quad n \in \mathbf{Z}$
(ii) $\mathbb{E}X_n = 0$ $\quad n \in \mathbf{Z}$
(iii) $\mathbb{E}[X_{n+m}\overline{X}_m]$ does not depend on m.

By Herglotz's lemma, we may associate with the covariance function

$$K_X(n) = \mathbb{E}[X_{m+n}\overline{X}_m] \qquad n \in \mathbf{Z}$$

a finite measure μ_X on $(]-\pi,+\pi], \mathscr{B}(]-\pi,+\pi]))$ such that for every n in $\mathbf{Z}$,

$$K_X(n) = \int e^{i\lambda n}\mu_X(d\lambda).$$

The measure μ_X is called the spectral measure of X, its density, if any, the spectral density of X.

Theorem B.8.2 Let $X = (X_n)_{n\in\mathbf{Z}}$ be a wide sense stationary sequence with spectral measure μ_X. There exists a unique OSM Z in $(]-\pi,+\pi], \mathscr{B}(]-\pi,+\pi]))$ with associated measure μ_X such that for every n in $\mathbf{Z}$,

$$X_n = \int e^{i\lambda n}Z(d\lambda).$$

We consider the OSM Z as defined on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ and carried by $]-\pi,+\pi]$, and put

$$X_t = \int e^{i\lambda t}Z(d\lambda) \qquad t \in \mathbb{R}.$$

The stationary process $(X_t)_{t\in\mathbb{R}}$ extends the stationary sequence $(X_n)_{n\in\mathbf{Z}}$ and is known as soon as the sequence (X_n) is known. Actually, in the general

case of a spectral measure with compact support, the following sampling statement holds.

Theorem B.8.3 Let $X = (X_t)_{t\in\mathbb{R}}$ be an L^2-continuous wide sense stationary process. Suppose that the spectral measure μ_X is absolutely continuous with bounded density and vanishes outside the interval $]-k,+k]$. Then, in $L^2(\Omega,\mathscr{A},\mathbb{P})$,

$$X_t = \lim_{N\uparrow\infty} \sum_{n=-N}^{+N} X_{n\pi/k} \frac{\sin(tk-n\pi)}{tk-n\pi}, \qquad t\in\mathbb{R}.$$

Proof Let Z_X be the OSM associated with X. Then

$$X_t = \int e^{i\lambda t} Z_X(d\lambda), \qquad t\in\mathbb{R}.$$

In the Hilbert space $L^2(]-k,+k], d\lambda)$, the Fourier representation of the function

$$e_t(\lambda) = e^{i\lambda t}$$

with respect to the orthonormal basis

$$g_n(\lambda) = \frac{1}{\sqrt{2k}} e^{i\lambda n\pi/k}, \qquad n\in\mathbf{Z},$$

is given by

$$\begin{aligned} e_t &= \sum_{n\in\mathbf{Z}} \langle e_t, g_n\rangle g_n \\ &= \sum_{n\in\mathbf{Z}} \sqrt{2k}\, \frac{\sin(tk-n\pi)}{tk-n\pi}\, g_n. \end{aligned}$$

As the density of μ_X is bounded and vanishes outside $]-k,+k]$, this expansion holds still in $L^2(\mathbb{R},\mu_X)$. It follows from the isometry property that

$$X_t = \lim_{N\uparrow\infty} \sum_{n=-N}^{+N} \frac{\sin(tk - n\pi)}{tk - n\pi} \int e^{i\lambda n\pi/k} Z_X(d\lambda)$$

$$= \lim_{N\uparrow\infty} \sum_{n=-N}^{+N} \frac{\sin(tk - n\pi)}{tk - n\pi} X_{n\pi/k}$$

in $L^2(\Omega, \mathscr{A}, \mathbb{P})$. □

If now the stationary process $X = (X_t)$ has a spectral measure with bounded density but no compact support, we put, for $k \in \mathbf{N}^*$,

$$\tilde{X}_t^{(k)} = \int e^{i\lambda t} 1_{]-k,+k]}(\lambda) Z_X(d\lambda)$$

and obtain

$$\tilde{X}_t^{(k)} = \lim_{n\uparrow\infty} \sum_{n=-N}^{+N} \tilde{X}_{n\pi/k}^{(k)} \frac{\sin(tk - n\pi)}{tk - n\pi}.$$

The quadratic error is given by

$$\mathbb{E}|X_t - \tilde{X}_t^{(k)}|^2 = \mu_X(\mathbb{R}\backslash]-k, +k]).$$

(b) Approximate Solutions of Differential Equations

Let U be a white noise (see Example B.4.2) with associated σ-finite OSM Z. Consider the pth-order ($p \geq 1$) linear differential equation

$$(*) \qquad a_0 X_t + a_1 \frac{dX_t}{dt} + \cdots + a_p \frac{d^p X_t}{dt^p} = U_t.$$

It can be shown that if the polynomial $P(x) = a_0 + a_1 x + \cdots + a_p x^p$ does not vanish on the imaginary axis, the equation (*) has a unique stationary solution, which is given by

$$X_t = \int \frac{e^{i\lambda t}}{P(i\lambda)} Z(d\lambda).$$

B.8.4 Exercise

Verify the above assertion.

Our aim is to find some approximation of X_t that can be simulated. Our method is to replace derivatives by difference quotients, and higher derivatives by higher difference quotients:

$$\frac{dX_t}{dt} \approx \frac{1}{h}(X_t - X_{t-h})$$

$$\vdots$$

$$\frac{d^pX_t}{dt^p} \approx \frac{1}{h^p}(X_t - C_p^1 X_{t-h} + \cdots + (-1)^k C_p^k X_{t-kh} + \cdots + (-1)^p X_{t-ph}).$$

The white noise is replaced by a discrete time white noise with variance $2\pi/h$. Indeed, for $h > 0$, we put

$$W_n = \sqrt{\frac{h}{2\pi}} \int e^{i\lambda nh} 1_{]-\pi/h,+\pi/h]}(\lambda) Z(d\lambda), \qquad n \in \mathbf{Z}.$$

The sequence $(W_n)_{n\in\mathbf{Z}}$ is centered orthonormal and the differential equation is replaced by the sequence of equations

$$\tilde{X}_{nh} P\left(\frac{1}{h}\right) + \cdots + (-1)^k \tilde{X}_{(n-k)h} \frac{1}{k!h^k} P^{(k)}\left(\frac{1}{h}\right) + \cdots + (-1)^p \frac{1}{h^p} \tilde{X}_{(n-p)h} a_p = \sqrt{\frac{2\pi}{h}} W_n, \qquad n \in \mathbf{Z}.$$

The sequence $(\tilde{X}_{nh})_{n\in\mathbf{Z}}$ is described by an autoregressive model of order q, and

(**) $$\tilde{X}_{nh} = b_0 W_n - b_1 \tilde{X}_{(n-1)h} - \cdots - b_p \tilde{X}_{(n-p)h}$$

with

$$b_0 = \frac{\sqrt{2\pi}}{\sqrt{h} P\left(\frac{1}{h}\right)}, \qquad b_k = (-1)^k \frac{P^{(k)}\left(\frac{1}{h}\right)}{k!h^k P\left(\frac{1}{h}\right)} \qquad k = 1, \ldots, p.$$

For the recurrence formula (**) to be useful in simulation, it must rapidly wipe out the p first arbitrary values $\tilde{X}_0, \ldots, \tilde{X}_{(p-1)h}$. Therefore the linear recurrence

$$f(n) = -b_1 f(n-1) - \cdots - b_p f(n-p)$$

has to be stable. Looking for solutions $f(n) = u^n, u \in \mathbf{C}$, we obtain by Taylor's formula

$$P\left(\frac{1}{h} - \frac{1}{uh}\right) = 0.$$

Suppose now that the roots of P lie in the half-plane $\operatorname{Re} z < 0$. Then $|u| < 1$ and the recursion is stable. The wide sense stationary sequence $(\tilde{X}_{nh})_{n \in \mathbf{Z}}$ is an autoregressive process with spectral density

$$f(\lambda) = \frac{1}{2\pi |Q(e^{-i\lambda})|^2} \, 1_{]-\pi,+\pi]}(\lambda),$$

where

$$Q(z) = \frac{1}{b_0}\,(1 + b_1 z + \cdots + b_p z^p).$$

Note that

$$Q(e^{-i\lambda}) = \frac{1}{b_0 P\left(\frac{1}{h}\right)} P\left(\frac{1}{h} - \frac{1}{he^{i\lambda}}\right) = \sqrt{\frac{h}{2\pi}}\, P\left(\frac{1 - e^{-i\lambda}}{h}\right),$$

and so

$$f(\lambda) = \frac{1}{h\left|P\left(\frac{1 - e^{-i\lambda}}{h}\right)\right|^2} \, 1_{]-\pi,+\pi]}(\lambda).$$

The covariance function is

$$\mathbb{E}[\tilde{X}_{(n+m)h}\overline{\tilde{X}}_{mh}] = \int e^{in\lambda} f(\lambda)\, d\lambda$$
$$= \int e^{in\lambda h} g_h(\lambda)\, d\lambda,$$

where

$$g_h(\lambda) = \frac{1}{\left| P\left(\frac{1 - e^{-i\lambda h}}{h} \right) \right|^2} 1_{]-\pi/h, +\pi/h]}(\lambda).$$

As $h \to 0$, $g_h(\lambda)$ converges uniformly and in L^1-norm to the spectral density of X, which we have seen to be

$$g(\lambda) = \frac{1}{|P(i\lambda)|^2}.$$

So the process

$$\tilde{X}_t = \int \frac{e^{i\lambda t}}{P\left(\frac{1 - e^{-i\lambda h}}{h} \right)} 1_{]-\pi/h, +\pi/h]}(\lambda) Z(d\lambda)$$

approximates process X_t; precise estimates can be derived from

$$X_t - \tilde{X}_t = \int e^{i\lambda t} \left(\frac{1}{P(i\lambda)} - \frac{1}{P\left(\frac{1 - e^{-i\lambda h}}{h} \right)} 1_{]-\pi/h, +\pi/h]}(\lambda) \right) Z(d\lambda).$$

B.9 Spectrum Discretization

We now consider stationary processes with discrete spectra, which are in some sense approximations of a given process X.

For instance, given a small $h > 0$ and a sequence of (centered) orthonormal random variables (N_n), we put

$$\tilde{X}_t = \sum_{n \in \mathbf{Z}} N_n (\mu_X(](n - \tfrac{1}{2})h, (n + \tfrac{1}{2})h]))^{1/2} e^{inht}.$$

Then

$$\mu_{\tilde{X}} = \sum_{n \in \mathbf{Z}} \mu_X(](n - \tfrac{1}{2})h, (n + \tfrac{1}{2})h])\delta_{nh}$$

and as $h \to 0$, $\mu_{\tilde{X}}$ converges narrowly to μ_X and $K_{\tilde{X}}$ to K_X uniformly in every compact subset of $\mathbb{R}$.

This kind of discretization preserves the order of magnitude of the decay at infinity of the spectral measure, and thus the local regularity (or irregularity) of paths is not altered. In the L^2-sense, the C^k-differentiability is the same for X and $\tilde{X}$ because it depends only on the finiteness of the integral

$$\int |\lambda|^{2k} \mu_X(d\lambda).$$

As for the a.s. continuity of paths, it follows from Kolmogorov's theorem that a continuous version for (X_t) exists if there exist positive constants α, β, ϵ such that

$$\mathbb{E}|X_t - X_s|^\alpha \leq \beta |t - s|^{1+\epsilon}$$

for all $t, s \in \mathbb{R}$. Taking $\alpha = 2$, we get

$$\mathbb{E}|X_{t+u} - X_t|^2 = \int |e^{i\lambda u} - 1|^2 \mu_X(d\lambda)$$

and thus an L^2-continuously differentiable process has a continuous version.

B.9.1 Exercise

Let $h > 0$, $(N_n)_{n \in \mathbf{Z}}$ be a sequence of orthonormal random variables, and $(a_n)_{n \in \mathbf{Z}}$ be a sequence of complex numbers such that there exists $\alpha > 2$ with $\sum_{n \in \mathbf{Z}} |a_n|^2 |n|^\alpha < \infty$. Show that the series $\sum N_n a_n e^{inht}$ converges a.s. in $\mathscr{C}((2\pi/h)\mathbf{T})$. [*Hint*: Estimate $\mathbb{P}(|N_k| > \sqrt{|k|} \log |k|)$ and use the Borel-Cantelli lemma.]

B.10 Simulating Stationary Processes with Stationary Germs

The simulation of stationary processes is often used to test the quality of indentification, filtering, or forecasting algorithms. It is therefore important for the simulated approximate process to have suitable ergodic properties.

We assume that X is a real Gaussian process:

$$X_t = \int_0^\infty \sqrt{2}\cos\,\lambda t\, f(\lambda) dB_\lambda^{(1)} + \int_0^\infty \sqrt{2}\sin\lambda t\, f(\lambda) dB_\lambda^{(2)},$$

where f is an even smooth function, and $B^{(1)}$ and $B^{(2)}$ are two independent Brownian motions. Its spectral density is $(f(\lambda))^2$. The following ideas would apply equally well, mutatis mutandis, to Gaussian stationary fields.

B.10.1 Zeroth-Order Germs

Let $(U_n)_{n\in\mathbf{Z}}$ be a sequence of independent standard normal variables. For $l > 0$, put

$$\psi_{n,l}(t) = \frac{\sin\left(\dfrac{tl}{2} - n\pi\right)}{\dfrac{tl}{2} - n\pi}, \qquad n \in \mathbf{Z}$$

(*) $$X_t^{(0)} = \sum_{-\infty}^{+\infty} U_n \psi_{n,l}(t).$$

As we have seen in the proof of Theorem B.8.3, the sequence of functions

$$f_n(\lambda) = \frac{1}{\sqrt{l}}\, e^{2i\pi n\lambda/l}$$

is an orthonormal basis in $H = L^2(]-l/2, +l/2[, d\lambda)$, and the Fourier coefficients of the function

$$e_t(\lambda) = e^{i\lambda t}$$

are given by

$$\langle e_t, f_n\rangle = \sqrt{l}\ \psi_{n,l}(t).$$

It follows that, for $s, t \in \mathbb{R}$,

$$\mathbb{E}[X^{(0)}_{t+s}X^{(0)}_s] = \sum_{-\infty}^{+\infty} \psi_{n,l}(t+s)\psi_{n,l}(s)$$
$$= \frac{1}{l}\,\langle e_{t+s}, e_s\rangle$$
$$= \frac{1}{l}\int_{-l/2}^{l/2} e^{i\lambda t}\,d\lambda.$$

The process $X^{(0)}$ is therefore stationary and its spectral density is

$$r(\lambda) = \frac{1}{l}\;1_{]-l/2,l/2[}(\lambda).$$

Conversely, if r is the spectral density of a real Gaussian stationary process X, the representation (*) holds with $U_n = X_{2n\pi/l}, n \in \mathbf{Z}$.

For $\theta \in \mathbb{R}$ and $a \in \mathbb{R}$, let us put

$$X^{(\theta,0)}_t = \sum_{-\infty}^{+\infty} a\sqrt{2l}\;\psi_{n,l}(t)(U_n \cos\,\theta t + V_n \sin\,\theta t),$$

where (U_n) and (V_n) are two independent sequences of independent standard normal variables. We have

$$\mathbb{E}[X^{(\theta,0)}_{t+s}X^{(\theta,0)}_s] = 2a^2 l \sum_{-\infty}^{+\infty} \psi_{n,l}(t+s)\psi_{n,l}(s)\cos\,\theta t$$
$$= 2a^2 \cos\,\theta t \int_{-(l/2)}^{l/2} e^{it\lambda}d\lambda,$$

which, for $|\theta| > l/2$, can be written as

$$a^2\left[\int_{-\theta-l/2}^{-\theta+l/2} e^{it\lambda}d\lambda + \int_{\theta-l/2}^{\theta+l/2} e^{it\lambda}d\lambda\right],$$

and for $\theta = 0$ can be written as

$$2a^2 \int_{-(l/2)}^{l/2} e^{it\lambda} d\lambda.$$

We conclude that for $k \in \mathbf{Z}$, the process $X^{(kl,0)}$ is stationary with spectral density

$$a^2(1_{]-|k|l-l/2,-|k|l+l/2[}(\lambda) + 1_{]+|k|l-l/2,+|k|l+l/2[}(\lambda))$$

for $k \neq 0$, and

$$2a^2 1_{]-l/2,l/2[}(\lambda)$$

for $k = 0$.

B.10.2 First-Order Germs

Again, let (U_n) and (V_n) be two independent sequences of independent standard normal variables. For $a_0, a_1 \in \mathbb{R}$ and $\theta \in \mathbb{R}$, put

$$X_t^{(\theta,1)} = \sqrt{2l} \sum_{-\infty}^{+\infty} [U_n(a_0 \cos\, \theta t \psi_{n,l}(t) + a_1 \sin\, \theta t \psi'_{n,l}(t)) + V_n(a_0 \sin\, \theta t \psi_{n,l}(t) - a_1 \cos\, \theta_t \psi'_{n,l}(t)],$$

where

$$\psi'_{n,l}(t) = \frac{1}{\sqrt{l}} \langle e'_t, g_n \rangle$$

with $e'_t(\lambda) = i\lambda e^{i\lambda t}$. Then, for $s, t \in \mathbb{R}$,

$$\mathbb{E}[X_{t+s}^{(\theta,1)} X_s^{(\theta,1)}]$$

$$= 2l[a_0^2 \cos\, \theta t \sum_{-\infty}^{+\infty} \psi_{n,l}(t+s)\psi_{n,l}(s) + a_1^2 \cos\, \theta t \sum_{-\infty}^{+\infty} \psi'_{n,l}(t+s)\psi'_{n,l}(s)$$

$$+ a_0 a_1 \sin\, \theta t \sum_{-\infty}^{+\infty} (\psi'_{n,l}(t+s)\psi_{n,l}(s) - \psi_{n,l}(t+s)\psi'_{n,l}(s)) =$$

$$= 2a_0^2 \cos\,\theta t\langle e_{t+s}, e_s\rangle + 2a_1^2 \cos\,\theta t\langle e'_{t+s}, e'_s\rangle + 4a_0a_1 \sin\,\theta t\langle e'_{t+s}, e_s\rangle$$
$$= \int_{-(l/2)}^{l/2} e^{i\lambda t}[a_0^2(e^{i\theta t} + e^{-i\theta t}) + a_1^2\lambda^2(e^{i\theta t} + e^{-i\theta t})$$
$$+ 2a_0a_1\lambda(e^{i\theta t} - e^{-i\theta t})]\,d\lambda$$
$$= \int_{-(l/2)}^{l/2} e^{i(\lambda+\theta)t}(a_0 + a_1\lambda)^2 d\lambda + \int_{-(l/2)}^{l/2} e^{i(\lambda-\theta)t}(a_0 - a_1\lambda)^2 d\lambda.$$

We conclude similarly that for $k \in \mathbf{Z}$, the process $X^{(kl,1)}$ is stationary with spectral density

$$(a_0 - a_1(\lambda + kl))^2 1_{]-kl-l/2,-kl+l/2[}(\lambda) + (a_0 + a_1(\lambda - kl))^2 1_{]kl-l/2,kl+l/2[}(\lambda)$$

for $k \neq 0$, and

$$2(a_0 + a_1\lambda)^2 1_{]-l/2,l/2[}(\lambda)$$

for $k = 0$.

B.10.3 Approximation

Let us take two independent sequences $(U_{n,k})$ and $(V_{n,k}), n \in \mathbf{Z}, k \in \mathbf{Z}$, of independent standard normal variables. As a first-order approximation of the process X, we consider the process

$$\tilde{X}_t^{l,0} = \sqrt{l} \sum_{k=-\infty}^{+\infty} f(kl) \sum_{n=-\infty}^{+\infty} \psi_{n,l}(t)(U_{n,k} \cos\, klt + V_{n,k} \sin\, klt).$$

This process is a sum of independent processes that are of the form studied in Section B.10.1 and whose spectral density we know. Hence $\tilde{X}^{l,0}$ is stationary and its spectral density is

$$\tilde{r}_0(\lambda) = \sum_{k=-\infty}^{+\infty} f^2(kl) 1_{]kl-l/2,kl+l/2[}(\lambda).$$

Similarly, using first-order germs, we get the approximation

$$\tilde{X}_t^{l,1} = \sqrt{l} \sum_{k=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} [U_{n,k}(f(kl) \cos klt\psi_{n,l}(t) + f'(kl) \sin klt\psi'_{n,l}(t))$$
$$+ V_{n,k}(f(kl) \sin klt\psi_{n,l}(t) - f'(kl) \cos klt\psi'_{n,l}(t))],$$

which is a stationary process with spectral density

$$\tilde{r}_1(\lambda) = \sum_{k=-\infty}^{+\infty} (f(kl) + (\lambda - kl)f'(kl))^2 1_{]kl-l/2,kl+l/2[}(\lambda).$$

Thus we have constructed stationary processes $\tilde{X}^{l,0}$ and $\tilde{X}^{l,1}$ whose spectral density is close to that of X. To study the quality of such an approximation, we suppose that the process X and the processes $\tilde{X}^{l,0}$ and $\tilde{X}^{l,1}$ are defined on the same probability space, and obtain L^2-estimates from which follow results in law that make sense here in a simulation framework. We proceed as follows.

The spectral representation of X is

$$X_t = \int_0^\infty \sqrt{2} \cos \lambda t f(\lambda) dB_\lambda^{(1)} + \int_0^\infty \sqrt{2} \sin \lambda t f(\lambda) dB_\lambda^{(2)},$$

so let us define $\tilde{f}_0$ and $\tilde{f}_1$ by

$$\tilde{f}_0(\lambda) = \sum_{k=-\infty}^{+\infty} f(kl) 1_{]kl-l/2,kl+l/2[}(\lambda)$$

and

$$\tilde{f}_1(\lambda) = \sum_{k=-\infty}^{+\infty} (f(kl) + (\lambda - kl)f'(kl)) 1_{]kl-l/2,kl+l/2[}(\lambda).$$

The functions $\tilde{f}_0$ and $\tilde{f}_1$ are even up to a negligible set, so if we put

$$Y_t^0 = \int_0^\infty \sqrt{2} \cos \lambda t \tilde{f}_0(\lambda) \, dB_\lambda^{(1)} + \int_0^\infty \sqrt{2} \sin \lambda t \tilde{f}_0(\lambda) \, dB_\lambda^{(2)}$$

and

$$Y_t^1 = \int_0^\infty \sqrt{2}\cos \lambda t \tilde{f}_1(\lambda)\, dB_\lambda^{(1)} + \int_0^\infty \sqrt{2}\sin \lambda t \tilde{f}_1(\lambda)\, dB_\lambda^{(2)}$$

then the processes Y^0 and Y^1 are stationary Gaussian and centered with density $(\tilde{f}_0)^2$ and $(\tilde{f}_1)^2$ respectively. So Y^0 has the same law as $\tilde{X}^{l,0}$ and Y^1 has the same law as $\tilde{X}^{l,1}$.

But we obviously have

$$\mathbb{E}|X_t - Y_t^0|^2 = 2\int_0^\infty |f(\lambda) - \tilde{f}_0(\lambda)|^2\, d\lambda$$

and

$$\mathbb{E}|X_t - Y_t^1|^2 = 2\int_0^\infty |f(\lambda) - \tilde{f}_1(\lambda)|^2\, d\lambda$$

B.10.4 Remarks

(1) In practice we have to truncate the infinite sums and approximate $\tilde{X}_t^{l,0}$ by

$$\tilde{\tilde{X}}_t^{l,0} = \sqrt{l} \sum_{k=-K}^{+K} f(kl) \sum_{n:|(tl/2)-n\pi|\le L} \psi_{n,l}(t)(U_{n,k}\cos klt + V_{n,k}\sin klt).$$

Then

$$\mathbb{E}|\tilde{X}_t^{l,0} - \tilde{\tilde{X}}_t^{l,0}|^2 = \sum_{|k|>K} f^2(kl)l + \sum_{k=-K}^{K} f^2(kl)l \sum_{n:|t(l/2)-n\pi|>L} \psi_{n,l}^2(t)$$

$$\le \sum_{|k|>K} f^2(kl)l + \sum_{k=-K}^{K} f^2(kl)l \frac{2}{\pi^2} \sum_{m=[L/\pi]}^{+\infty} \frac{1}{m^2},$$

which can be made small by choosing K and L sufficiently large.

(2) The processes $\tilde{X}^{l,0}$ and $\tilde{X}^{l,1}$ are ergodic and mixing since they are Gaussian with a spectral density [Cornfeld, Fomin, and Sinai, 1982].

(3) Naturally, we can repeat what we have done with zeroth-order and first-order germs with pth-order germs as well.

B.10.5 Remarks

Nonlinear Representations
Iterative definitions such as

$$X_{n+1} = F(X_n, X_{n-1}, \cdots, X_{n-p}, U_n)$$

where F is a nonlinear function and (U_n) is a discrete time Gaussian white noise, may give rise to strictly stationary processes. Such processes are as easy to simulate as AR-processes, but their theoretical study is more delicate (they are generally non-Gaussian and their second-order structure does not define them in law); see for example Guégan [1988].

Representations by Wiener Chaos
In some models, it is interesting to extend the particular features of Gaussian stationary processes to non-Gaussian processes.

This can be done with the help of the decomposition in chaos of the L^2-space of the Brownian motion which was discovered by Wiener: Let $\Omega = \mathscr{C}(\mathbb{R}_+, \mathbb{R})$ equipped with its Borel σ-field β, and let m be the Wiener measure on (Ω, β). The coordinates $(B_t)_{t\geq 0}$ are a standard Brownian motion process under m. Every random variable $X \in L^2(\Omega, \beta, m)$ can be written uniquely as

$$X = \sum_{n=0}^{\infty} \int_{0<t_1<\cdots<t_n} f_n(t_1, \cdots, t_n) dB_{t_1} \cdots dB_{t_n},$$

where $f_n \in L^2(P_n, \lambda_n)$,

$$P_n = \{(t_1, \ldots, t_n) : 0 < t_1 < \cdots < t_n\},$$

and λ_n is the Lebesgue measure on P_n.

Then by the isometry property of the stochastic integral (e.g., see Bouleau and Hirsch [1991])

$$\mathbb{E}X^2 = \sum_{n=0}^{\infty} \| f_n \|^2_{L^2(P_n, \lambda_n)}.$$

By this isometry, every unitary group of operators on the space $\oplus_{n=0}^{\infty} L^2(P_n, \lambda_n)$ (which is called the Fock space) defines a wide sense stationary process.

For simplicity, we consider the second chaos. Suppose that Q_t is a group of unitary operators on $L^2(P_2, \lambda_2)$ (we can take, for example, $Q_t f(t_1, t_2) = e^{it(t_1}f(t_1, t_2)$). The process

$$X_t = \int_{0<t_1<t_2} Q_t f(t_1, t_2) dB_{t_1} dB_{t_2}$$

is a wide sense stationary process.

If, moreover, we take for Q_t the translation along the diagonal of P_2, we obtain

$$X_t = \int_{0<t_1<t_2} f(t + t_1, t + t_2) dB_{t_1} dB_{t_2}$$

with $f \in L^2(P_2, \lambda_2)$, which is a strictly stationary process. (The process is defined here only for $t \geq 0$, but that could be altered by a slight change of setting.)

C MARKOV PROCESSES

This section is devoted to the simulation of Markov chains and Markov processes with transition probability functions that are given by an analytic expression or can, at least, be simulated.

We shall deal only with time homogeneous $\mathbb{R}^d$-valued Markov processes for which there exists a version with right-continuous paths. Typically, if $(P_t)_{t\geq 0}$ is the semi-group of linear operators on $\mathbb{R}^d$ that generates the Markov process, the canonical version will be denoted by

$$(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, \sigma_t, \mathbb{P}_\nu),$$

where

$$\Omega = \{\text{right-continuous mappings from } \mathbb{R}_+ \text{ into } \mathbb{R}^d\}$$
$$X_t(\omega) = \omega(t) \qquad \text{(coordinate mappings)}$$
$$\mathcal{F}_t = \sigma(X_s, s \leq t) \qquad t \geq 0$$
$$\mathcal{F} = \vee_t \mathcal{F}_t$$
$$\sigma_t(\omega)(s) = \omega(t + s) \qquad \text{(shift operators)} \quad s, t \geq 0$$

ν = probability measure on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ (initial distribution)
$\mathbb{P}_\nu$ = probability measure on $(\Omega, \mathscr{F})$ such that $\mathbb{P}_\nu \circ X_0^{-1} = \nu$

and

$$\mathbb{E}_\nu[f(X_t)|X_u, u \le s] = P_{t-s}f(X_s) \qquad \mathbb{P}_\nu - a.s., \qquad 0 \le s < t < \infty.$$

C.1 Jump Processes

For sake of brevity, we write $(E, \mathscr{E})$ for $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. Consider a Markov kernel P of $(E, \mathscr{E})$ to itself and a nonnegative bounded Borel function λ on $(E, \mathscr{E})$. Let M_λ be the operator on $B\mathscr{E}$ defined by

$$M_\lambda f = \lambda f, \qquad f \in b\mathscr{E}$$

and consider the semi-group

$$P_t = e^{tM_\lambda(P-I)},$$

where the exponential is understood in $\mathscr{L}(b\mathscr{E}, b\mathscr{E})$. This semi-group defines a Markov process, which is called a Markov jump process. It behaves like a Markov chain that has kernel P, but whose sojourn time at $x \in E$ is exponentially distributed with parameter $\lambda(x)$.

More precisely, let $(Z_n)_{n\ge0}$ be a Markov chain on E with initial distribution ν and transition kernel P. Let $(U_n)_{n\ge1}$ be a Borel sequence in [0, 1] independent of $(Z_n)_{n\ge0}$ and, for $n \ge 1$, put

$$T_n(Z_{n-1}) = -\frac{1}{\lambda(Z_{n-1})} \log(U_n) \qquad \text{if} \quad \lambda(Z_{n-1}) > 0$$
$$= +\infty \qquad \text{if} \quad \lambda(Z_{n-1}) = 0.$$

Then, for $t \ge 0$, the process

$$(*) \qquad X_t = Z_0 1_{\{t<T_1(Z_0)\}} + \sum_{n=1}^{\infty} Z_n 1_{\{\sum_{k=1}^{n} T_k(Z_{k-1}) \le t < \sum_{k=1}^{n+1} T_k(Z_{k-1})\}}$$

is a Markov jump process with semi-group $(P_t)_{t\ge0}$.

C.1.1 Exercise

Prove the last assertion.

It is clear that (*) provides a method for simulating Markov jump processes. For instance, let $(V_n)_{n\geq 0}$ be a sequence of i.i.d. random variables with common distribution ν and let $(N_t)_{t\geq 0}$ be a Poisson process with parameter λ independent of $(V_n)_{n\geq 0}$. Put

$$X_t = \sum_{n=0}^{N_t} V_n, \qquad t \geq 0$$

Then X is a Markov jump process with

$$Pf(x) = \int f(x+y)\, d\nu(y)$$
$$\lambda(x) = \lambda, \qquad \forall x \in E.$$

C.1.2 Exercise

Show that in this case X has independent stationary increments.

C.2 Iterative or Recursive Simulation

> Die Frage bei allem, was du tun willst: "ist es so, dass ich es unzählige Male tun will?" ist das grösste Schwergewicht.[2] [F. Nietzsche, Leipzig, 1901].

The first method (iterative) consists of a time discretization: we get a Markov chain $(X_{nh})_{n\geq 0}$ with transition kernel P_h in place of $(X_t)_{t\geq 0}$. Thus we obtain a sequence of points on a trajectory but have no way to get information about the behavior of the trajectory between these points. This method is useful for studying the asymptotic behavior.

The second method (recursive) involves recursive interpolation between the times s and t by computing the conditional distribution of X_u given $X_s = a$ and $X_t = b$ for $s < u < t$. (See Figure 3.3.)

The following five graphs are obtained by recursive simulation of a one-dimensional Brownian motion. The Brownian path is looked at with an increasing scale, which shows its fractal structure.

[2]The question which is the main center of gravity about anything that you want to do: "do I want to do it infinitely often in this way?"

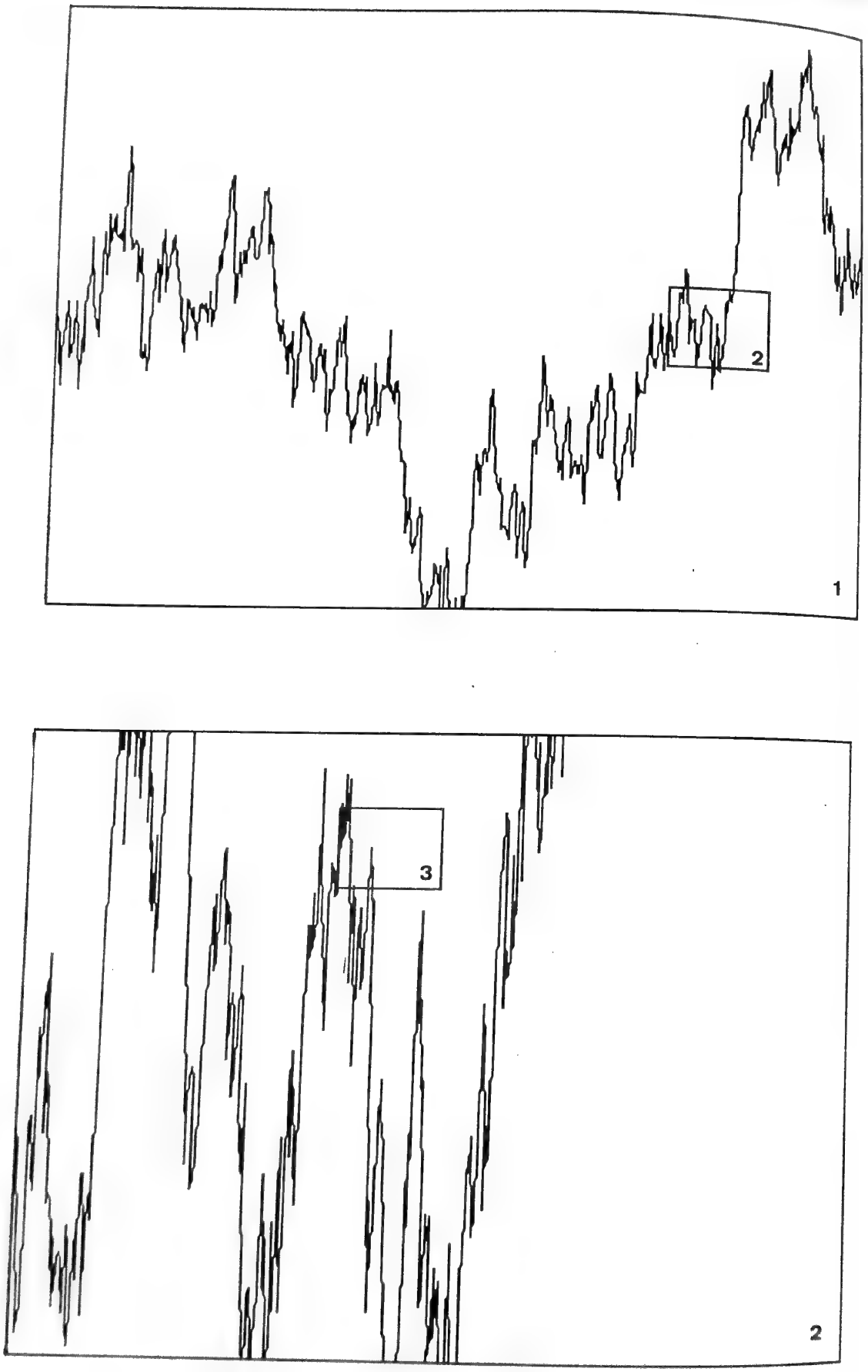


Figure 3.3

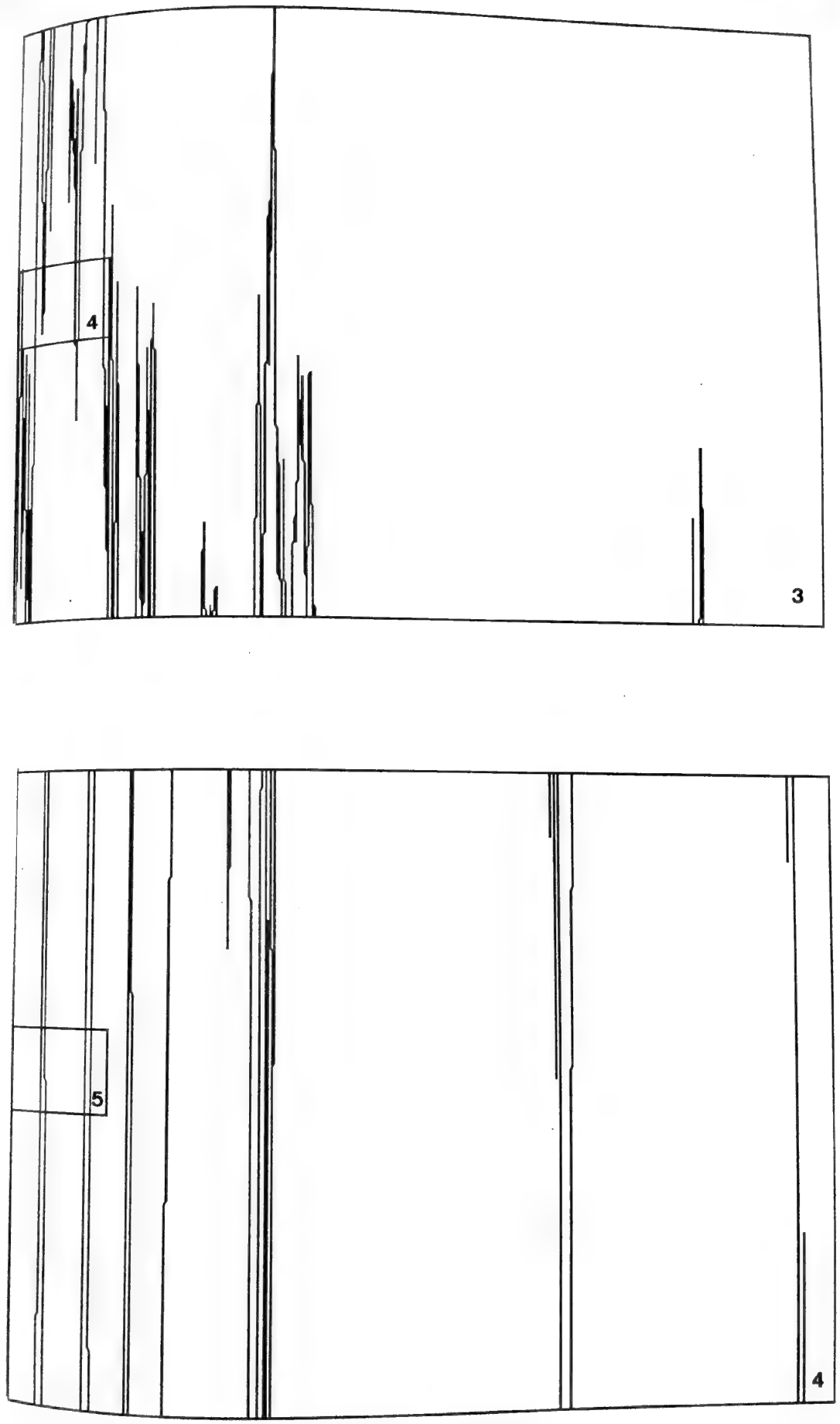


Figure 3.3 (*Continued*)

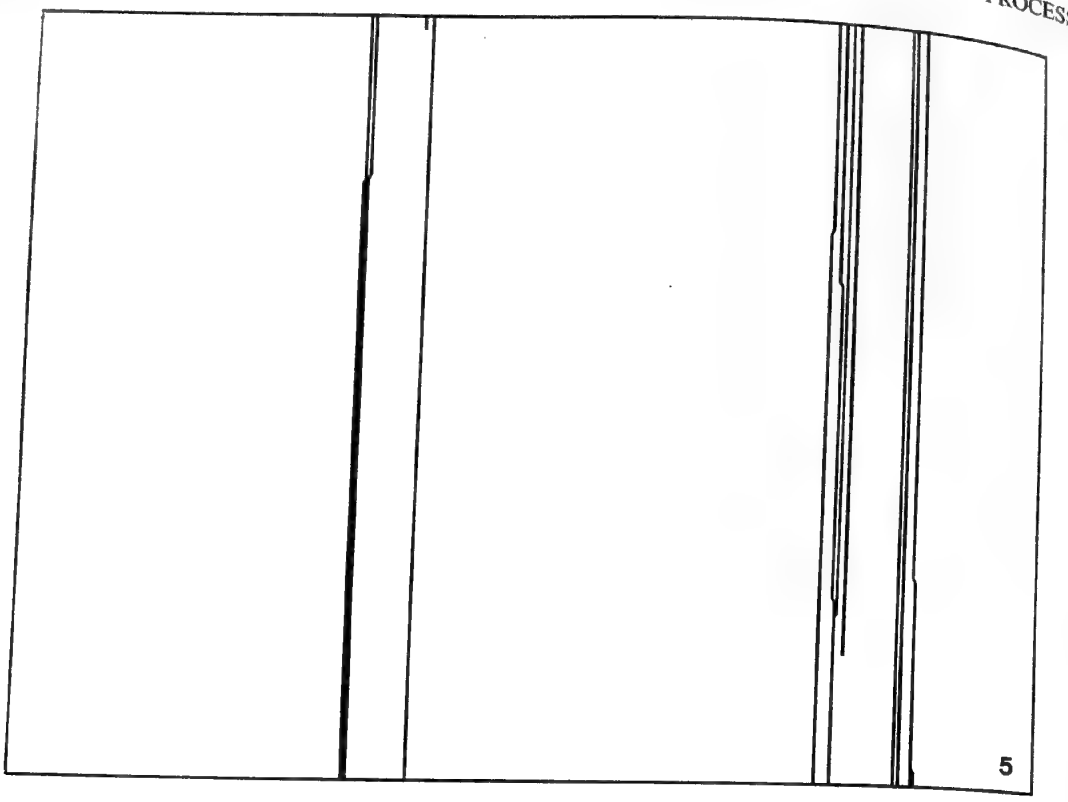


Figure 3.3 (*Continued*)

C.2.1 Exercise

Suppose that the transition probability function satisfies $P_t(x, dy) = p(t, x, y)\,dy$. Show that, for $0 \leq s < u < t$, there exists a regular conditional probability distribution of X_u given $(X_s = a, X_t = b)$, which is given by

$$\begin{cases} \dfrac{p(u-s,a,x)p(t-u,x,b)}{p(t-s,a,b)}\,dx & \text{if} \quad p(t-s,a,b) > 0 \\ \delta_a & \text{if} \quad p(t-s,a,b) = 0. \end{cases}$$

Je pense d'ailleurs que la manière dont, dans des exposés plus récents j'ai défini la fonction $X(t)$ du mouvement brownien est plus claire que celle de Wiener. Je vais la rappeler brièvement. Supposant pour fixer les idées $X(0) = 0$, on se donne une suite de variables gaussiennes réduites ξ_n, indépendantes les unes des autres; chaque ξ_n sera l'accroissement de $X(t)$ dans l'intervalle $(n-1, n)$ de sorte que

$$X(n) = \xi_1 + \xi_2 + \cdots + \xi_n$$

est bien de la forme $\xi\sqrt{n}$, la lettre ξ désignant toujours une variable gaussienne réduite. Il reste à définir la variation de $X(t)$ dans chacun des intervalles $(n-1, n)$.

Il n'y a pour cela qu'à appliquer la formule d'interpolation

$$(*) \qquad X\left(t+\frac{\tau}{2}\right)=\frac{X(t)+X(t+\tau)}{2}+\frac{\xi'}{2}\sqrt{\tau}$$

[...] Par application successive de la formule (*), on détermine $X(t)$ pour des valeurs de t multiples impaires de $\frac{1}{2}$, puis pour ceux de $\frac{1}{4}$, puis pour ceux de $\frac{1}{8}$, et ainsi de suite. Désignons par $X_p(t)$ la fonction égale si t est un multiple de 2^{-p} aux valeurs ainsi obtenues et variant linéairement dans chacun des intervalles qui séparent ces multiples. On démontre aisément que presque sûrement, pour p infini, $X_p(t)$ tend vers une limite $X(t)$, et cela uniformément dans tout intervalle fini. [...]

On arrive d'ailleurs à d'autres définitions constructives de la même fonction aléatoire en choisissant autrement les points de division sur l'axe des t. Il faudrait seulement utiliser une formule d'interpolation plus générale que (*).

Il faut bien comprendre le rôle de cette formule. Il serait facile de définir directement chacune des fonctions $X_p(t)$ indépendamment des autres, les $X_p(h/2^p)(h = 1,2,\ldots)$ s'obtenant successivement par l'addition de variables indépendantes de la forme $2^{-(p/2)}\xi$. Mais il n'y aurait aucune corrélation entre les différentes approximations et on ne pourrait plus parler de convergence vers une limite bien définie. Il n'y aurait que convergence en loi.[3] [Paul Lévy, 1970]

[3]Furthermore, the way in which, in more recent expositions, I have defined the function $X(t)$ of Brownian motion is clearer than that of Wiener. I will summarize it briefly. Suppose that $X(0) = 0$ and we are given a sequence of reduced Gaussian variables, independent from each other; each ξ_n is the amount of $X(t)$ increases in the interval $(n-1,n)$, so that

$$X(n)=\xi_1+\xi_2+\cdots+\xi_n$$

is of the form $\xi\sqrt{n}$, the letter ξ always denoting a reduced Gaussian variable. We must still define the way $X(t)$ varies in each of the intervals $(n-1,n)$. For this, we need only apply the interpolation formula

$$(*) \qquad X\left(t+\frac{\tau}{2}\right)=\frac{X(t)+X(t+\tau)}{2}+\frac{\xi'}{2}\sqrt{\tau}$$

By successive applications of (*), we determine $X(t)$ for values of t equal to odd multiples of $\frac{1}{2}$, then of $\frac{1}{4}$, then of $\frac{1}{8}$, and so on. Let us write $X_p(t)$ for the function that, when t is a multiple of 2^{-p}, equals the values we have obtained, and varies linearly in each of the intervals that separates these multiples. It is easily shown that, as p tends to infinity, $X_p(t)$ tends almost surely to a limit $X(t)$, and that this is uniform in every finite interval. We could moreover get other constructive definitions of the same random function by choosing other points dividing the t axis. We would need only to use a more general interpolation formula than (*). It is worthwhile understanding the role of this formula. It would be easy to define

C.2.2 Example

Let $(X_t)_{t\geq 0}$ be a subordinator with explicit transition probability density $p(t,x,y) = q(t,y-x)$. We give two examples:

- Γ-process,

$$q(t,x) = \frac{1}{\Gamma(t)}\, x^{t-1}e^{-x}, \qquad x > 0;$$

- the one-sided stable process of order $1/2$,

$$q(t,x) = \frac{1}{\sqrt{2\pi}}\, tx^{-3/2} \exp\left\{-\frac{t^2}{2x}\right\}, \qquad x > 0.$$

We compute the stopping time

$$T_a = \inf\{t > 0 : X_t \geq a\}, \qquad a > 0,$$

as follows:

(a) the number

$$N = \inf\{n > 0 : X_{nh} \geq a\},$$

where h is fixed, is computed by iterative simulation;

(b) between $(N-1)h$ and Nh we search by binary division the interval where X_t crosses the threshold a. Thus T_a can be computed for each trajectory with a priori fixed accuracy.

C.2.3 Example

Let $(X_t^1)_{t\geq 0}$ and $(X_t^2)_{t\geq 0}$ be two independent subordinators from the above class. The subordinator defined by

---

each of the functions $X_p(t)$ independently from the others, the $X_p(h/2^p)(h = 1, 2, \ldots)$ being given successively by adding independent variables of the form $2^{-p/2}\xi$. But there would be no correlation between the different approximations, and we could not speak of convergence to a well-defined limit. There would only be convergence in law.

$$Y_t = X^1_{X^2_t}, \qquad t \geq 0$$

has transition probability density

$$p(t,x,y) = \int q_1(z, y-x) q_2(t,z)\, dz$$

but the use of this formula requires rather lengthy and imprecise quadratures. It is more convenient to proceed as follows. First, by iterative simulation, compute

$$N = \inf\{n > 0 : X^1_{X^2_{nh}} \geq a\}$$

and store $(X^2_{(N-1)h}, X^2_{Nh})$ as well as $(X^1_{X^2_{(N-1)h}}, X^1_{X^2_{Nh}})$. By recursive computation,

(a) simulate $X^2_{(N-1)h+(h/2)}$ given $X^2_{(N-1)h}$ and X^2_{Nh};
(b) simulate $X^1_{X^2_{(N-1)h+(h/2)}}$ given $X^1_{X^2_{(N-1)h}}$ and $X^1_{X^2_{Nh}}$

and systematically divide the time interval where Y crosses the level a. In this way T_a is computed with a priori fixed accuracy.

C.3 Simulation of a Brownian Motion Until Hitting a Hyperplane

We now turn to the so-called importance sampling method in particle transport problems. This method has close connections with the theory of Doob h-processes.

We use the following principle: in order to compute $\mathbb{E}_x F(X_t) = P_t F(x)$ for a Markov process (X_t) we consider a positive Borel function φ and define the semi-group

$$Q_t(f) = \frac{1}{\varphi} P_t(\varphi f).$$

If φ is (P_t)-excessive, this semi-group is sub-Markovian and defines a new Markov process Y_t. We compute

$$P_t(F)(x) = \varphi(x) Q_t\left(\frac{F}{\varphi}\right)(x)$$

by simulating Y_t. This method may have the advantage of lowering the variance of the random variable whose expectation is computed.

As an illustration, we consider simulating an $\mathbb{R}^d$-valued Brownian motion ($d \geq 2$) until it hits the hyperplane $H = \{x_d = 0\}$. Let

$$\tau = \inf\{t \geq 0 : B_t \in H\}.$$

If h is the Poisson kernel in the half-space, the h-transform of the Brownian motion yields a new Markov process that has the same distribution as the Brownian motion conditioned to converge to 0 (or any other point in H) at time τ and then to be killed.

We now proceed to explain how this can be done. We first need some classical results about one-dimensional Brownian motion. These are well-known, but we give short proofs for the sake of completeness.

Proposition C.3.1 Let (B_t) be a one-dimensional Brownian motion starting from 0, and let $A > 0$. We define

$$T_a = \inf\{t \geq 0 : B_t = a\}.$$

(i) (Reflection principle). Put

$$\begin{aligned} B'_t &= B_t && \text{if} \quad t \leq T_a \\ &= 2a - B_t && \text{if} \quad t \geq T_a. \end{aligned}$$

Then (B'_t) is also a Brownian motion starting from 0.

(ii) The distribution of T_a is given by

$$\mathbb{P}_0(T_a \leq t) = \int_0^t \frac{a}{\sqrt{2\pi s^3}} \exp\left\{-\frac{a^2}{2s}\right\} ds, \qquad t \geq 0.$$

Proof (i) Clearly B' starts from 0 and has continuous paths. Intuitively, both $B_{t+T_a} - a$ and $a - B_{t+T_a}$ are Brownian motions independent of the σ-field $\mathcal{F}_{T_a}$. More rigorously, let $0 \leq s < t < \infty$, $A \in \mathcal{F}_s$ and u be a real number. It is easily checked that

$$\mathbb{E}_0[\exp\{iu(B'_t - B'_s)\} 1_{\{T_a \leq s\}} 1_A] = \mathbb{E}_0[\exp\{iu(B_t - B_s) 1_{\{T_a \leq s\}} 1_A]$$

and

$$\mathbb{E}_0[\exp\{iu(B'_t - B'_s)\}1_{\{t<T_a\}}1_A] = \mathbb{E}_0[\exp\{iu(B_t - B_s)1_{\{t<T_a\}}1_A].$$

It follows from the optional sampling theorem for martingales that

$$\mathbb{E}_0\left[\exp\left\{\pm iuB_t + \frac{u^2}{2}\,t\right\}\middle|\mathscr{F}_{t\wedge T_a}\right] = \exp\left\{\pm iuB_{t\wedge T_a} + \frac{u^2}{2}(t\wedge T_a)\right\}$$

and we deduce that

$$\begin{aligned}
&\mathbb{E}_0[\exp\{iu(B'_t - B'_s)\}1_{\{s<T_a\le t\}}1_A] \\
&\quad = \mathbb{E}_0[\exp\{iu(2a - B_s)\}\mathbb{E}_0[\exp\{-iuB_t\}|\mathscr{F}_{t\wedge T_a}]1_{\{s<T_a\le t\}}1_A] \\
&\quad = \mathbb{E}_0[\exp\{-iuB_s\}1_{\{s<T_a\le t\}}\mathbb{E}_0[\exp\{iuB_t\}|\mathscr{F}_{t\wedge T_a}]1] \\
&\quad = \mathbb{E}_0[\exp\{iu(B_t - B_s)\}1_{\{s<T_a\le t\}}1_A].
\end{aligned}$$

Summing the three equalities we get

$$\mathbb{E}_0[\exp\{iu(B'_t - B'_s)\}1_A] = \mathbb{P}_0(A)\exp\left\{-\frac{u^2}{2}\,(t-s)\right\}$$

which proves that $B'_t - B'_s$ is independent of $\mathscr{F}_s$ and has distribution $N(0, t-s)$.

(ii)

$$\begin{aligned}
\{T_a \le t\} &= \{T_a \le t, B_t \ge a\} \cup \{T_a \le t, B_t < a\} \\
&= \{T_a \le t, B_t \ge a\} \cup \{T_a \le t, B'_t > a\} \\
&= \{B_t \ge a\} \cup \{B'_t > a\}.
\end{aligned}$$

These events are disjoint, so

$$\begin{aligned}
\mathbb{P}_0(T_a \le t) &= \mathbb{P}_0(B_t \ge a) + \mathbb{P}_0(B'_t > a) \\
&= \frac{2}{\sqrt{2\pi t}}\int_a^\infty e^{-(x^2/2t)}\,dx.
\end{aligned}$$

Finally, we make a change of variables $x = (t^{1/2}a)s^{-1/2}$ to obtain

$$\mathbb{P}_0(T_a \le t) = \int_0^t (2\pi s^3)^{-1/2} a e^{-(a^2/2s)}\, ds.$$

□

Proposition C.3.2 Let $x > 0$, let B_t be a one-dimensional Brownian motion starting from x, and let

$$T_0 = \inf\{t \ge 0 : B_t = 0\}.$$

Then, for any $b > 0$,

$$\mathbb{P}_x(B_t \le b, T_0 > t) = \int_0^b \pi(t, x, y)\, dy,$$

where

$$\pi(t, x, y) = p(t, x, y) - p(t, x, -y)$$

and

$$p(t, x, y) = (2\pi t)^{-1/2} \exp\{-(y - x)^2/2t\}.$$

Proof We have

$$\mathbb{P}_x(B_t \le b, T_0 > t) = \mathbb{P}_x(0 \le B_t \le b) - \mathbb{P}_x(0 \le B_t \le b, T_0 \le t).$$

Put

$$\begin{aligned} B'_t &= B_t \quad \text{if} \quad t \le T_0 \\ &= -B_t \quad \text{if} \quad t \ge T_0. \end{aligned}$$

It follows from the reflection principle that B' is a Brownian motion starting from x and

$$\begin{aligned} \mathbb{P}_x(B_t \le b, T_0 > t) &= \mathbb{P}_x(0 \le B_t \le b) - \mathbb{P}_x(-b \le B'_t \le 0) \\ &= \int_0^b (p(t, x, y) - p(t, x, -y))\, dy. \end{aligned}$$

□

We return to a d-dimensional Brownian motion B. Let E_+ be the upper half-space in $\mathbb{R}^d$ and let H be the boundary hyperplane, that is,

$$E_+ = \{z = (x, y) \in \mathbb{R}^d; x \in \mathbb{R}^{d-1}, y > 0\},$$
$$H = \{z = (x, y) \in \mathbb{R}^d; x \in \mathbb{R}^{d-1}, y = 0\}.$$

Put
$$\tau = \inf\{t > 0 : B_t \notin E_+\}.$$

We are interested in the behavior of the Brownian motion killed as it leaves E_+. Suppose $z_0 = (x_0, y_0) \in E_+$ and set

$$b_t = (B_t^1, \ldots, B_t^{d-1}), \qquad t \geq 0.$$

Since the exit time τ depends only on the last coordinate, it follows from Proposition C.3.1 that

$$\mathbb{P}_{z_0}(\tau \leq t) = \int_0^t \frac{y_0}{(2\pi s^3)^{1/2}} \exp\left\{-\frac{y_0^2}{2s}\right\} ds$$

and the conditional probability density of b_τ given $\tau = t_0$ is

$$g(t_0, z_0, \theta) = (2\pi t_0)^{-(d-1)/2} \exp\left\{-\frac{|\theta - x_0|^2}{2t_0}\right\}, \qquad \theta \in \mathbb{R}^{d-1}.$$

These expressions provide an approach to simulating τ and B_τ:

(a) first pick τ with probability density

$$l(t, z_0) = (2\pi t^3)^{-1/2} y_0 \exp\{-y_0^2/2t\}, \qquad t > 0;$$

(b) once we have computed $\tau = t_0$, pick b_τ with density $g(t_0, z_0, \theta)$.

All that remains is to simulate the process B between time 0, when it lies at z_0, and time t_0 when it reaches H at $\overline{\theta_0} = (\theta_0, 0)$. This can be done recursively, for which purpose we must first compute some conditional distributions.

Proposition C.3.3 Let $0 \leq t_1 < s < t_2 < t_0$. A regular version of the conditional probability density of B_s given $\tau = t_0, B_\tau = \theta_0, B_{t_1} = z_1$, and $B_{t_2} = z_2$ (θ_0 in $\mathbb{R}^{d-1}$, z_1 and z_2 in E_+) is given by

$$\frac{\pi(s-t_1,z_1,z)\pi(t_2-s,z,z_2)}{\pi(t_2-t_1,z_1,z_2)}, \qquad z \in E_+,$$

where

$$\pi(t,(x,y),(x',y')) = (2\pi t)^{-d/2} \exp\left\{-\frac{|x'-x|^2}{2t}\right\} \cdot \left(\exp\left\{-\frac{(y'-y)^2}{2t}\right\} - \exp\left\{-\frac{(y'+y)^2}{2t}\right\}\right).$$

Proof Let $\varphi_1, \psi, \varphi_2$ be bounded Borel functions on E_+, k be a bounded Borel function on $\mathbb{R}^{d-1}$, and ρ be a bounded Borel function on $\mathbb{R}_+$, which vanishes outside $]t_2, +\infty[$. From the Markov property of Brownian motion, it follows that, provided $t_1 > 0$,

$$\begin{aligned}
&\mathbb{E}_{z_0}[\varphi_1(B_{t_1})\psi(B_s)\varphi_2(B_{t_2})\rho(\tau)k(b_\tau)] \\
&\quad = \mathbb{E}_{z_0}[\varphi_1(B_{t_1})1_{\{\tau>t_1\}}\psi(B_s)1_{\{\tau>s\}}\varphi_2(B_{t_2})1_{\{\tau>t_2\}}\rho(\tau)k(b_\tau)] \\
&\quad = \mathbb{E}_{z_0}[\varphi_1(B_{t_1})1_{\{\tau>t_1\}}\mathbb{E}_{B_{t_1}}[\psi(B_{s-t_1})1_{\{\tau>s-t_1\}} \\
&\qquad \mathbb{E}_{B_{s-t_1}}[\varphi_2(B_{t_2-s})1_{\{\tau>t_2-s\}}\mathbb{E}_{B_{t_2-s}}[\rho(t_2+\tau)k(b_\tau)]]]] \\
&\quad = \int_{E_+} \pi(t_1,z_0,z_1)\varphi_1(z_1)\,dz_1 \int_{E_+} \pi(s-t_1,z_1,z)\psi(z)\,dz \\
&\qquad \int_{E_+} \pi(t_2-s,z,z_2)\varphi_2(z_2)\,dz_2 \\
&\qquad \int_{t_2}^{\infty} \rho(t_0)l(t_0-t_2,z_2)\,dt_0 \int_{\mathbb{R}^{d-1}} k(\theta_0)g(t_0-t_2,z_2,\theta_0)\,d\theta_0,
\end{aligned}$$

while

$$\begin{aligned}
&\mathbb{E}_{z_0}[\varphi_1(B_{t_1})\varphi_2(B_{t_2})\rho(\tau)k(b_\tau)] \\
&\quad = \int_{E_+} \pi(t_1,z_0,z_1)\varphi_1(z_1)dz_1 \int_{E_+} \pi(t_2-t_1,z_1,z_2)\varphi_2(z_2)\,dz_2 \\
&\qquad \int_{t_2}^{\infty} \rho(t_0)l(t_0-t_2,z_2)\,dt_0 \int_{\mathbb{R}^{d-1}} k(\theta_0)g(t_0-t_2,z_2,\theta_0)d\theta_0.
\end{aligned}$$

The conclusion follows from the definition of the conditional distribution. In the case $t_1 = 0$, we remove the first left integral; nothing else is altered. □

The case $t_2 = t_0$ requires a separate computation.

Proposition C.3.4 Let $0 \le t_1 < s < t_0$. A regular version of the conditional probability density of B_s given $\tau = t_0, b_\tau = \theta_0, B_{t_1} = z_1 (\theta_0$ in $\mathbb{R}^{d-1}, z_1$ in E_+) is

$$\frac{\pi(s - t_1, z_1, z) l(t_0 - s, z) g(t_0 - s, z, \theta_0)}{l(t_0 - t_1, z_1) g(t_0 - t_1, z_1, \theta_0)}, \qquad z \in E_+.$$

Proof Let φ_1, ψ, k, and ρ be as in the proof of Proposition C.3.3, except that ρ now vanishes outside $]s, +\infty[$ instead of $]t_2, +\infty[$. We check that, for $t_1 > 0$,

$$\begin{aligned}
&\mathbb{E}_{z_0}[\varphi_1(B_{t_1})\psi(B_s)\rho(\tau)k(b_\tau)] \\
&\quad = \int_{E_+} \pi(t_1, z_0, z_1)\varphi_1(z_1)dz_1 \int_{E_+} \pi(s - t_1, z_1, z)\psi(z)\, dz \\
&\quad \int_s^\infty l(t_0 - s, z)\rho(t_0)dt_0 \int_{\mathbb{R}_{d-1}} g(t_0 - s, z, \theta_0)k(\theta_0)\, d\theta_0
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}_{z_0}[\varphi_1(B_{t_1})\rho(\tau)k(b_\tau)] \\
&\quad = \int_{E_+} \pi(t_1, z_0, z_1)\varphi_1(z_1)dz_1 \int_s^\infty l(t_0 - t_1, z_1)\rho(t_0)\, dt_0 \\
&\quad \int_{\mathbb{R}_{d-1}} g(t_0 - t_1, z_1, \theta_0)k(\theta_0)\, d\theta_0.
\end{aligned}$$

This gives the conclusion, which is trivially the same in the case $t_1 = 0$. □

We can now proceed: we pick successively the exit time t_0, the exit place θ_0, and some positions for $B_{t_1}, \ldots, B_{t_n}$ if $0 < t_1 < \cdots < t_n < t_0$. In some particular instances, we may only be interested in trajectories that hit a given subset A of H. For that purpose, we use a rejection method in the choice of B_τ given $\tau = t_0$. It is, however, of some interest to compute the global distribution of B_τ, which we now do.

Proposition C.3.5 Let A be a Borel set in $\mathbb{R}^{d-1}$. Then

$$\mathbb{P}_z(B_\tau \in A) = \int_A h(\theta, z)\, d\theta, \qquad z \in \mathbb{E}_+,$$

where h is the Poisson kernel of the half-space, that is,

$$h(\theta, z) = \frac{\Gamma(d/2)}{\pi^{d/2}} \frac{y}{|\bar{\theta} - z|^d}, \qquad \theta \in \mathbb{R}^{d-1}, \quad z \in E_+.$$

Proof

$$\begin{aligned}\mathbb{P}_z(B_\tau \in A) &= \int_0^\infty \int_A g(t, z, \theta) l(t, z)\, dt\, d\theta \\ &= \int_A \int_0^\infty \frac{y}{(2\pi)^{d/2} t^{d/2+1}} \exp\{-|\bar{\theta} - z|^2/2t\}\, dt\, d\theta.\end{aligned}$$

A change of variables $s = |\bar{\theta} - z|^2/2t$ gives

$$\mathbb{P}_z(B_\tau \in A) = \int_A \int_0^\infty \frac{y}{(2\pi)^{d/2}} \frac{2^{d/2}}{|\bar{\theta} - z|^d} s^{(d-1)/2} e^{-s} ds\, d\theta$$

and integrating over $]0, \infty[$ gives the required result. □

C.3.6 Remark

We first draw the hitting time $\tau = t_0$: in this way the Poisson kernel h does not appear in the simulation formulas. In other instances, it is more convenient to draw the hitting place on the boundary first and then to handle the process conditioned to hit the boundary at this place, which is exactly the philosophy of Doob's h-transforms. Let us give some details. For θ in $\mathbb{R}^{d-1}$, z in E_+, φ in $b(\mathscr{B}(E_+))$, put

$$\begin{aligned} h_\theta(z) &= h(\theta, z), \\ \pi_t \varphi(z) &= \int_{E_+} \pi(t, z, z') \varphi(z')\, dz' = \mathbb{E}_z[\varphi(B_t) 1_{\{\tau > t\}}], \qquad t \geq 0, \\ Q_t^\theta \varphi(z) &= \frac{1}{h_\theta(z)} \pi_t(h_\theta \varphi)(z), \qquad t \geq 0. \end{aligned}$$

We check that for s, t in $\mathbb{R}_+$,

$$\begin{aligned}\pi_{t+s}\varphi(z) &= \mathbb{E}_z[\varphi(B_{t+s})1_{\{\tau>t+s\}}] \\ &= \mathbb{E}_z[\varphi(B_s\circ\sigma_t)1_{\{\tau>t\}}1_{\{\tau\circ\sigma_t>s\}}] \\ &= \mathbb{E}_z[1_{\{\tau>t\}}\mathbb{E}_{B_t}[\varphi(B_s)1_{\{\tau>s\}}]] \\ &= \pi_t\pi_s\varphi(z).\end{aligned}$$

This proves that $(\pi_t)_{t\geq 0}$ is a sub Markovian semi-group of linear operators on $b(\mathscr{B}(E_+))$. A straightforward calculation yields the same conclusion for $(Q_t^\theta)_{t\geq 0}$, for any θ in $\mathbb{R}^{d-1}$. Then, for $\varphi_1,\ldots,\varphi_n$ in $b(\mathscr{B}(E_+))$ and k a bounded Borel function on $\mathbb{R}^{d-1}$,

$$\begin{aligned}&\mathbb{E}_z[\varphi_1(B_{t_1})\cdots\varphi_n(B_{t_n})1_{\{\tau>t_n\}}k(b_\tau)] \\ &\quad= \int_{E_+}\pi(t_1,z,z_1)\varphi_1(z_1)\,dz_1\cdots\int_{E_+}\pi(t_n-t_{n-1},z_{n-1},z_n) \\ &\qquad\cdot\varphi_n(z_n)\,dz_n\int_{\mathbb{R}^{d-1}}h(\theta,z_n)k(\theta)\,d\theta \\ &\quad= \int_{\mathbb{R}^{d-1}}h(\theta,z)k(\theta)\,d\theta\,\frac{1}{h_\theta(z)}\int_{E_+}\pi(t_1,z,z_1)\varphi_1(z_1) \\ &\qquad\cdot h_\theta(z_1)\,\frac{dz_1}{h_\theta(z_1)}\cdots\int_{E_+}\pi(t_n-t_{n-1},z_{n-1},z_n)\varphi_n(z_n)h_\theta(z_n)\,dz_n.\end{aligned}$$

We can deduce that, for almost every θ in $\mathbb{R}^{d-1}$,

$$\begin{aligned}&\mathbb{E}_z[\varphi_1(B_{t_1})\cdots\varphi_n(B_{t_n})1_{\{\tau>t_n\}}|b_\tau=\theta] \\ &\qquad= Q_{t_1}^\theta[\varphi_1 Q_{t_2-t_1}^\theta[\cdots Q_{t_n-t_{n-1}}^\theta[\varphi_n]\cdots](z).\end{aligned}$$

This equality gives a precise meaning to our statement above: the h_θ-transform $(Q_t^\theta)_{t\geq 0}$ of the semi-group $(\pi_t)_{t\geq 0}$ of the Brownian motion killed at the boundary H is the semi-group of the Brownian motion conditioned to hit the boundary at place θ and be immediately killed.

C.3.7 Remark

To deal with balls instead of half-spaces we use a Kelvin transformation that maps a ball one-to-one onto the half-space E_+. In particular, the exit time from a ball of a Brownian motion starting from its center has the explicit Laplace transform

$$\mathbb{E}_0 e^{-pT} = \frac{(p/2)^{(d-2)/4}}{\Gamma(d/2)I_{d/2-1}(\sqrt{2p})},$$

where I_ν is the modified Bessel function

$$I_\nu(z) = \sum_{m=0}^{\infty} \frac{(z/2)^{\nu+2m}}{m!\Gamma(\nu+m-1)}, \qquad z > 0.$$

We refer to Durrett [1984], Knight [1981], and Ciesielski and Taylor [1962] for further information.

D PROCESSES WITH STATIONARY INDEPENDENT INCREMENTS

Among $\mathbb{R}^d$-valued Markov processes, the processes with stationary independent increments (PSII) deserve a particular study because of their analytical properties and their role in the construction of models. We have already introduced these processes in Section B.7, Definitions B.7.4 and B.7.5. In the present section, we study them in more detail and concentrate on their application in simulation. We shall examine a number of topics: the relationship between the independent increments property and the fact that the transition semi-group commutes with space translations (convolution semi-groups); the Brownian motion as a PSII; the increasing PSIIs or subordinators; the $\mathbb{R}^d$-valued PSII and the stable PSII; the notion of subordination in Bochner sense and its interest for the simulation of infinitely divisible laws; and the existence of a realization of all subordinated processes of a general Markov process.

D.1 Definition and First Properties

We adopt the following definition, which extends Definition B.7.4 by removing assumptions (i) and (ii) (for a definition with property (iii) below weakened, see, for example, Bretagnolle [1973]).

Definition D.1.1 A process with stationary independent increments (PSII) is a process $(X_t)_{t\geq 0}$ with values in $\mathbb{R}^d$ such that

- **(i)** $\forall s, t, 0 \leq s < t, X_t - X_s$ is independent of the σ-field $\mathcal{F}_s = \sigma(X_u, u \leq s)$;
- **(ii)** the law of $X_t - X_s$ depends only on $t - s$;
- **(iii)** the sample paths of X are almost surely continuous on the right with left limits.

This implies that X is a homogeneous Markov process. Indeed, let $\mu_{t-s}(dx)$ be the law of $X_t - X_s$. For $u \in \mathbb{R}^d$,

$$\mathbb{E}[e^{i\langle u, X_t\rangle}|\mathscr{F}_s] = \mathbb{E}[e^{i\langle u, X_t - X_s + X_s\rangle}|\mathscr{F}_s]$$
$$= e^{i\langle u, X_s\rangle}\int e^{i\langle u,y\rangle}\mu_{t-s}(dy) = \int e^{i\langle u, y+X_s\rangle}\mu_{t-s}(dy).$$

Thus if a sub-Markov kernel P_t is defined on $(\mathbb{R}, \mathscr{B}(\mathbb{R}^d))$ by

$$P_t f(x) = \int f(x+y)\mu_t(dy), \qquad \forall f \in b.\mathscr{B}(RR^d), \tag{D.1.2}$$

the family $(P_t)_{t\geq 0}$ is a semi-group and

$$\mathbb{E}[e^{i\langle u, X_t\rangle}|\mathscr{F}_s] = (P_{t-s}e^{i\langle u,.\rangle})(X_s), \qquad \forall u \in \mathbb{R}^d.$$

This relation extends, since linear combinations of exponential functions form an algebra generating the Borel σ-field, into

$$\mathbb{E}[f(X_t)|\mathscr{F}_s] = P_{t-s}f(X_s), \qquad \forall f \in b.\mathscr{B}(\mathbb{R}^d). \tag{D.1.3}$$

Moreover, if f is bounded and continuous, the relation

$$\int f(x)\mu_t(dx) = \mathbb{E}[f(X_t - X_0)]$$

and the right continuity of the paths give

$$\lim_{t\to 0}\int f(x)\mu_t(dx) = f(0).$$

We have the following proposition.

Proposition D.1.4 A PSII X is a Markov process admitting a transition semi-group (P_t) given by

$$P_t f(x) = \int f(x+y)\mu_t(dy), \qquad \forall f \in b.\mathscr{B}(\mathbb{R}^d),$$

where μ_t is the law of $X_t - X_0$. The family of probability measures (μ_t) is a convolution semi-group on $\mathbb{R}^d$:

$$\mu_t * \mu_s = \mu_{t+s}, \quad \forall s, t \in \mathbb{R}_+$$
$$\lim_{t \to} \mu_t = \delta_0 \quad \text{narrowly.}$$

Conversely, let X be an $\mathbb{R}^d$-valued Markov process whose sample paths are continuous on the right and with left limits and which admit a transition semi-group (Q_t) that commutes with translations of space:

$$\text{(D.1.5)} \qquad Q_t(\tau_x f) = \tau_x(Q_t f), \qquad \forall f \in b\mathscr{B}(\mathbb{R}^d) \quad \forall x \in \mathbb{R}^d,$$

where $\tau_x f$ denotes the function $y \to f(x + y)$. Then the mapping $f \in \mathscr{C}_0(\mathbb{R}^d) \to Q_t f(0)$ defines a Radon measure, and hence a probability measure ν_t and we have

$$Q_t f(x) = \int f(x + y)\nu_t(dy).$$

It follows that $\nu_t * \nu_s = \nu_{t+s}$ and

$$\mathbb{E}[f(X_t - X_s)|\mathscr{F}_s] = \mathbb{E}[f(X_t - X_s)] = \int f(y)\nu_{t-s}(dy).$$

Therefore X is a PSII in the sense of Definition D.1.1.

D.2 Convolution Semi-groups on $\mathbb{R}^d$

The families of probability measures on $\mathbb{R}^d$, which are a semi-group under the convolution product and are narrowly continuous (see Proposition D.1.4), have been characterized by Lévy and Hincin and are related to several analytical and probabilistic properties. We shall quote only the main results and refer to Feller [1966, Chapters IX and XVII] and to Berg and Forst [1975] for a more detailed exposition.

D.2.1 Negative Definite Functions

If X is a PSII, the characteristic function of $X_t - X_0$ is of the form

$$\text{(D.2.2)} \qquad \mathbb{E}[e^{i\langle u, X_t - X_0\rangle}] = e^{-t\psi(u)}, \qquad \forall t \in \mathbb{R}_+.$$

The functions ψ obtained in this way are the negative definite functions, which are continuous and vanish at zero. They satisfy

$$\text{(D.2.3)} \qquad \sum_{i,j=1}^{n} [\psi(u_i) + \overline{\psi(u_j)} - \psi(u_i - u_j)]\xi_i \overline{\xi_j} \geq 0$$

$\forall (u_1, \ldots, u_n) \in \mathbb{R}^n, \forall (\xi_1, \ldots, \xi_n) \in \mathbf{C}^n$.

D.2.4 Lévy–Hincin Formula

The continuous negative definite functions vanishing at zero can be represented in the form

$$\text{(D.2.5)} \qquad \psi(u) = i\langle \mu, u \rangle + \tfrac{1}{2}\langle \Sigma u, u \rangle + \int_{\mathbb{R}^d} [1 - e^{i\langle y,u \rangle} + i\langle y, u \rangle 1_{\{|y| \leq 1\}}]\, d\nu(y),$$

where $\mu \in \mathbb{R}^d$, Σ is a semi-definite positive matrix, and ν is a positive measure on $\mathbb{R}^d$ such that $\nu\{0\} = 0$ and

$$\int_{\mathbb{R}^d} 1 \wedge |y|^2 d\nu(y) < +\infty.$$

The measure ν is not in general σ-finite on $\mathbb{R}^d$, but only on $\mathbb{R}^d \backslash \{0\}$. It is uniquely determined by the knowledge of ψ. It also possesses a probabilistic interpretation with respect to the sample paths of the process as follows.

If A is a Borel set in $\mathbb{R}^d \backslash \{0\}$, $t\nu(A)$ is the expectation of the number of jumps on $]0, t]$ whose amplitude is in A:

$$t\nu(A) = \mathbb{E} \sum_{0 < s \leq t} 1_A(X_s - X_{s-}).$$

If formula (D.2.5) is changed by replacing the expression between the brackets by

$$\left[1 - e^{i\langle y,u \rangle} + i\langle y, u \rangle \frac{|y|}{1 + |y|} \right]$$

only the coefficient μ has to be changed; the formula is valid with the same measure ν. The measure ν is called the *Lévy measure* of the PSII X.

D.3 Convolution Semi-groups on $\mathbb{R}_+$

The families of probability measures on $\mathbb{R}_+$ which are a semi-group under the convolution product and are narrowly continuous, can be studied using the Laplace transform instead of the characteristic functions.

D.3.1 Bernstein Functions

If X is a real PSII with increasing sample paths (subordinator), then

$$\mathbb{E}[e^{-p(X_t - X_0)}] = e^{-t\varphi(p)}, \qquad \forall p \in \mathbb{R}_+. \tag{D.3.2}$$

The functions φ obtained in this way are the Bernstein functions vanishing at zero. These functions have remarkable properties:

- They are the functions f that are C^∞ from $]0, \infty[$ to $\mathbb{R}$ such that $f \geq 0, f(0) = 0$ and $(-1)^n (d^n f/dx^n) \leq 0, \forall n \geq 1$.
- They form a cone and are concave.
- They can be extended analytically to holomorphic functions on the half-space Re $z > 0$.
- If (φ_n) is a sequence of Bernstein functions that converges pointwise, the limit is Bernstein.
- They admit the integral representation

$$\varphi(x) = bx + \int_0^\infty (1 - e^{-xy})\, d\nu(y), \tag{D.3.3}$$

where $b \in \mathbb{R}_+$ and ν is a positive measure on $\mathbb{R}_+^*$ such that $\int_0^\infty (y \wedge 1)\, d\nu(y) < +\infty$. The number b and the measure ν are uniquely determined by φ. The measure ν is the Lévy measure of the subordinator associated with φ by (D.3.2).

- They operate on negative definite functions: if ψ is continuous negative definite vanishing at zero and φ is Bernstein vanishing at zero, then the function $\varphi \circ \psi$ is continuous negative definite vanishing at zero (see Sections D.5 and D.7 below).

- The cone of Bernstein functions vanishing at zero forms a monoid under composition: if φ_1 and φ_2 are Bernstein vanishing at zero, so is $\varphi_1 \circ \varphi_2$.

D.4 Examples

We give below some families of PSII whose transition semi-group has explicit forms. These PSII therefore can be simulated, either iteratively or recursively (see Section C.2 above). With the help of the subordination in the Bochner sense, they enable us to generate other simulatable families (see Section D.7 below).

D.4.1 Brownian Motion in $\mathbb{R}^d$

The negative definite function

$$u \in \mathbb{R}^d \to \tfrac{1}{2}|u|^2$$

corresponds to the heat semi-group given by

$$P_t f(x) = \frac{1}{(2\pi t)^{d/2}} \int e^{-(1/2t)|y|^2} f(x+y)\, dy$$

for $f \in (b\mathscr{B}(\mathbb{R}^d)$ (see Section C.3 above).

D.4.2 Cauchy Process in $\mathbb{R}^d$

The function $u \to |u|$ is negative definite and corresponds to the Cauchy semi-group given by

$$P_t f(x) = \Gamma\left(\frac{d+1}{2}\right) \int \frac{1}{t[\pi(|y|^2 + t^2)]^{(d+1)/2}} f(x+y)\, dy$$

(This is the heat semi-group subordinated by the stable subordinator of order $\frac{1}{2}$; see Section D.7.)

D.4.3 Γ-Process

The convolution semi-group (μ_t) on $\mathbb{R}_+$ given by

$$\mu_t(dx) = 1_{]0,\infty[}(x)\,\frac{1}{\Gamma(t)}\,x^{t-1}e^{-x}\,dx$$

defines a subordinator with Bernstein function

$$\varphi(p) = \log(1+p), \qquad p \in \mathbb{R}_+$$

and Lévy measure

$$\nu(dy) = 1_{]0,\infty[}(y)\,\frac{e^{-y}}{y}\,dy.$$

D.4.4 Poisson Process on $\mathbb{R}_+$

The convolution semi-group (μ_t) on $\mathbb{R}_+$ given by

$$\mu_t = \sum_{n=0}^{\infty} e^{-t}\frac{t^n}{n!}\,\delta_{na}$$

which correspond to the Lévy measure $\nu(dy) = \delta_a$, generates the Poisson process on $\mathbb{R}_+$.

D.4.5 Jump PSIIs

Let $U_1, U_2, \ldots, U_n, \ldots$ be a sequence of i.i.d. $\mathbb{R}^d$-valued random variables with common law μ, and let (N_t) be a Poisson process on $\mathbb{R}_+$ that vanishes at zero, has intensity $\lambda \in \mathbb{R}_+^*$, and is independent from the sequence (U_n).

Let

$$X_t = X_0 + \sum_{n=1}^{N_t} U_n,$$

with the convention $\sum_{n=1}^{0} = 0$. This defines a PSII with transition semi-group P_t given by

$$P_t = e^{\lambda t(P-I)},$$

where I is the identity, P is given by

$$Pf(x) = \int f(x+y)\,d\mu(y),$$

and the exponential is taken in the sense of bounded linear operators from $b\mathcal{B}(\mathbb{R}^d)$ into itself.

The associated negative definite function is given by

$$e^{-t\psi(u)} = \mathbb{E}e^{i\langle u, X_t - X_0\rangle} = \exp\left[\lambda t \int_{\mathbb{R}^d} (e^{i\langle u,x\rangle} - 1)\, d\mu(x)\right].$$

We see that the Lévy measure of X is the bounded measure $\lambda.\mu(dx)$.

We postpone to Section D.6 below the presentation of stable PSII. Before doing this we shall give a construction of a subordinator starting from its Lévy measure as an introduction to a probabilistic study of PSIIs.

D.5 Subordinators

Let φ be a Bernstein function with integral representation (see formula (D.3.3))

$$\varphi(x) = \int_0^\infty (1 - e^{-xy})\, d\nu(y),$$

where ν is a positive measure on $\mathbb{R}_+^*$ with respect to which $y \to y \wedge 1$ is integrable. We wish to show the existence of an increasing PSII X such that

$$\mathbb{E}e^{-p(X_t - X_0)} = e^{-t\varphi(p)} \qquad p \geq 0.$$

(a) Suppose first that the total mass $\|\nu\|$ is finite. As in Section D.4.5, if $U_1, \ldots, U_n, \ldots$ is an i.i.d. sequence with law $\nu/\|\nu\|$ and (N_t) is a Poisson process on $\mathbb{R}_+$ with intensity $\|\nu\|$, independent of the sequence (U_n), then the process

$$X_t = X_0 + \sum_{n=1}^{N_t} U_n$$

is convenient.

(b) Let (Y_t^1) and (Y_t^2) be two independent PSIIs of this type, with bounded Lévy measures ν^1 and ν^2 on $\mathbb{R}_+$. It is easily seen that $Y_t = Y_t^1 + Y_t^2$ is a PSII and that the Laplace transform $\Phi_t(p) = \mathbb{E}e^{-p(X_t - X_0)}$ satisfies (with the obvious notation)

$$\Phi_t(p) = \Phi_t^1(p)\Phi_t^2(p)$$
$$= \exp\left\{-t\|\nu_1\|\int(1-e^{-px})\frac{d\nu_1(x)}{\|\nu_1\|} - t\|\nu_2\|\int(1-e^{-px})\frac{d\nu_2(x)}{\|\nu_2\|}\right\}$$

and so its Lévy measure is $\nu_1 + \nu_2$.

Like Y^1 and Y^2, the process Y has only finitely many jumps on every finite interval. We quote the fact that if

$$Y_t^1 = Y_0^1 + \sum_{n=1}^{N_t^1} U_n^1$$

and

$$Y_t^2 = Y_0^2 + \sum_{n=1}^{N_t^2} U_n^2$$

are the relations of the constructions of Y^1 and Y^2, and the measures ν^1 and ν^2 are carried by disjoint Borel sets A^1 and A^2 in $\mathbb{R}_+\backslash\{0\}$, then

- $Y_t^1 - Y_0^1$ is the sum of the jumps of Y on $]0,t]$ whose amplitude is in A^1
- $Y_t^2 - Y_0^2$ is the sum of the jumps of Y on $]0,t]$ whose amplitude is in A^2.

(c) Now let ν be a given positive measure on $\mathbb{R}_+^*$ integrating $x \to x \wedge 1$, not necessarily bounded. We define bounded measures $\nu^{(n)}$ as follows:

$$\forall n \geq 1 \qquad \nu^{(n)}(dx) = 1_{]1/(n+1),1/n]}.\nu\ (dx)$$

and for $n = 0$,

$$\nu^{(0)}(dx) = 1_{]1,\infty[}.\nu\,(dx).$$

We consider a sequence $Y^{(n)}, n \geq 0$, of processes of the type above, associated with the measures $\nu^{(n)}$.

Put

$$X_t^{(n)} = \sum_{k=0}^{n} Y_t^{(k)}.$$

By an argument similar to that in (b), $(X_t^{(n)})$ is a PSII with Laplace transform

$$\Phi_t^{(n)}(p) = \mathbb{E}e^{-p(X_t^{(n)} - X_0^{(n)})}$$
$$= \exp\left[-t\int_{]1/(n+1),\infty[}(1 - e^{-px})\,d\nu(x)\right].$$

For fixed t, since $\int_0^\infty (x \wedge 1)\,d\nu(x)$ is finite, we have by the Lebesgue dominated convergence theorem that

$$\Phi_t^{(n)}(p) \to_{n\uparrow\infty} \Phi_t(p) = \exp\left[-t\int_{]0,\infty[}(1 - e^{-px})\,d\nu(x)\right].$$

As the sequence $X_t^{(n)}$ is increasing,

$$\lim_n X_t^{(n)} = \sum_{n=0}^{\infty} Y_t^{(n)} \qquad \text{a.s.}$$

and so the formula

$$X_t = \sum_{k=0}^{\infty} Y_t^{(k)}$$

defines a random variable with Laplace transform $\Phi_t(p)$.

To show that (X_t) is a PSII, we must show that its increments are independent, which follows easily from the fact that the $(Y_t^{(k)})$ are independent PSIIs, and we must show that (X_t) has sample paths continuous on the right with left limits, which follows from

(i) (X_t) has increasing sample paths since it is the limit of increasing processes,

(ii) by the dominated convergence theorem,

$$\lim_{s\downarrow t} X_s = \lim_{s\downarrow t} \sum_{k=0}^{\infty} Y_s^{(k)} = \sum_{k=0}^{\infty} \lim_{s\downarrow t} Y_s^{(k)}$$
$$= \sum_{k=0}^{\infty} Y_t^{(k)} = X_t.$$

D.5.1 Remark

Since the measures $\nu^{(k)}$ are carried by disjoint sets, the process $(Y_t^{(k)})$ represents the sum of the jumps of $X_s, s \le t$, with amplitude in $]1/(k+1), 1/k]$ (resp., in $]1, \infty[$ for $k = 0$). The average number of these jumps on $]0,1[$ is therefore

$$\|\nu^{(k)}\| = \nu\left(\left]\frac{1}{k+1}, \frac{1}{k}\right]\right) \qquad (\nu(]1,\infty[) \quad \text{if} \quad k = 0).$$

It follows that the expectation of the total number of jumps of X on $[0,1]$ is

$$\nu(]1,\infty[) + \sum_{k=1}^{\infty} \nu\left(\left]\frac{1}{k+1}, \frac{1}{k}\right]\right) = \|\nu\|.$$

Thus we see that if $\|\nu\| = +\infty$, the expectation of the number of jumps on $[0,1]$ is infinite. It can, moreover, be shown that there are infinitely many jumps a.s. on every nonempty open interval.

D.5.2 Example

One sided stable process of order $\frac{1}{2}$. By the equality (take the derivative of both sides)

$$\int_0^{\infty} (1 - e^{-px}) \frac{dx}{x^{3/2}} = \frac{1}{2}\Gamma\left(\frac{1}{2}\right)\sqrt{p} = \frac{\sqrt{\pi}}{2}\sqrt{p}, \qquad p > 0,$$

$(\sqrt{\pi}/2)\sqrt{p}$ is the Bernstein function associated with the measure $\nu = 1_{]0,\infty[}.(dx/x^{3/2})$.

Let X be the subordinator that corresponds to ν by the formula $\mathbb{E}e^{-p(X_t - X_0)} = e^{-t(\sqrt{\pi}/2)\sqrt{p}}$. The law μ_t of $X_t - X_0$ is

$$\mu_t(dx) = \frac{t}{4x^{3/2}}\, e^{-(\pi t^2/16x)} 1_{]0,\infty[}(x)\, dx.$$

D.6 Stable PSIIs

$\mathbb{R}^d$-valued PSII X is *stable* if $\forall c > 0, \exists a > 0, \exists b \in \mathbb{R}^d$ such that the process (cX_t) has the same law as the process $(X_{at} + bt)$. If we can take $b = 0$, X is said to be *strictly stable*. By the Lévy–Hincin formula it can be shown that we must have $a = c^\alpha$ for $\alpha \in]0, 2]$. The number α is the *order* of the stable PSII X.

D.6.1 Stable Laws on $\mathbb{R}$

In the real case, the negative definite function ψ of the Lévy–Hincin formula must have the form

$$\text{(D.6.2)} \qquad \begin{cases} \psi(u) = i\mu' u + c|u|^\alpha \left[1 - i\beta \operatorname{sign}(u) \tan\left(\alpha\frac{\pi}{2}\right)\right] \\ \text{for } \alpha \in]0, 2];\ \alpha \neq 1;\ c > 0;\ \beta, \mu' \in \mathbb{R}, |\beta| \leq 1. \end{cases}$$

$$\text{(D.6.3)} \qquad \begin{cases} \psi(u) = i\mu' u + c|u| \left[1 - i\beta \operatorname{sign}(u) \frac{2}{\pi} \log |u|\right] \\ \text{for } \alpha = 1, c > 0;\ \beta, \mu' \in \mathbb{R}, |\beta| \leq 1. \end{cases}$$

Among these laws, the strictly stable ones are those with

- $\mu' = 0$ in the case $\alpha \in]0, 2], \alpha \neq 1$,
- $\beta = 0$ in the case $\alpha = 1$.

In particular, the symmetric stable laws on $\mathbb{R}$ are obtained with $\psi(u) = c|u|^\alpha, \alpha \in]0, 2]$.

By (D.6.2) and (D.6.3) we can express the Lévy measure ν by:

$$\text{(D.6.4)} \qquad \nu(dx) = c\, \frac{\alpha}{\Gamma(2-\alpha)} \frac{1-\alpha}{\cos\frac{\pi\alpha}{2}} \frac{1}{|x|^{\alpha+1}} \cdot \left[\frac{1+\beta}{2}\, 1_{]0,+\infty[}(x) + \frac{1-\beta}{2}\, 1_{]-\infty,0[}(x)\right] dx$$

and, for $\alpha = 1$,

$$\text{(D.6.5)} \qquad \nu(dx) = c\,\frac{2}{\pi}\,\frac{1}{x^2}\left[\frac{1+\beta}{2}\,1_{]0,+\infty[}(x) + \frac{1-\beta}{2}\,1_{]-\infty,0[}(x)\right] dx.$$

If $\beta = 1$, the Lévy measure is carried by $\mathbb{R}_+^*$. These stable processes are called completely dissymmetric. If, in addition, $0 < \alpha < 1$ and $\mu' > 0$, then the process X is increasing and is a subordinator such that

$$\text{(D.6.6)} \qquad \mathbb{E}[e^{-p(X_t - X_0)}] = e^{-t\mu' p} \exp\left[-\frac{ctp^\alpha}{\cos\dfrac{\pi\alpha}{2}}\right].$$

D.6.7 Stable Laws on $\mathbb{R}^d$

The stable laws of order 2 are Gaussian. For $\alpha \in]0, 2[$, the negative definite function of the Lévy–Hincin formula must be of the form

$$\psi(u) = i\langle\mu, u\rangle + \int_{\mathbb{R}^d} [1 - e^{i\langle y,u\rangle} + i\langle y, u\rangle 1_{\{|y|\le 1\}}]\, d\nu(y),$$

where the Lévy measure ν is a product measure $\nu(dr, d\theta) = n(dr)m(d\theta)$, where m is a probability measure on the sphere S^{d-1} and n is a measure of the form $k(dr/r^{1+\alpha})$ on $]0, \infty[$.

We obtain the expressions

$$\text{(D.6.8)} \qquad \begin{cases} \psi(u) = i\langle\mu, u\rangle + c\displaystyle\int_{S^{d-1}} |\langle\theta, u\rangle|^\alpha \\ \quad \cdot\left[1 - i\,\text{sign}\,(\langle\theta, u\rangle)\tan\dfrac{\pi\alpha}{2}\right] m(d\theta) \\ \quad \text{for}\,\mu \in \mathbb{R}^d, c > 0, \alpha \in]0, 2[, \alpha \neq 1. \end{cases}$$

$$\text{(D.6.9)} \qquad \begin{cases} \psi(u) = i\langle\mu, u\rangle + c\displaystyle\int_{S^{d-1}} |\langle\theta, u\rangle| \\ \quad \cdot\left[1 + i\,\text{sign}\,(\langle\theta, u\rangle)\dfrac{2}{\pi}\log\,|\langle\theta, u\rangle|\right] m(d\theta) \\ \quad \text{for}\,\mu \in \mathbb{R}^d, \alpha = 1, c > 0. \end{cases}$$

The constant k of the Lévy measure is given by

$$k = c\,\frac{\alpha}{\Gamma(2-\alpha)}\,\frac{1-\alpha}{\cos\dfrac{\pi\alpha}{2}} \quad \text{for} \quad \alpha \neq 1$$

$$k = c\,\frac{2}{\pi} \quad \text{for} \quad \alpha = 1.$$

If $\mu = 0$ and the measure m is symmetric with respect to the origin, we obtain the stable symmetric laws on $\mathbb{R}^d$ with

(D.6.10) $$\psi(u) = c\int_{S^{d-1}} |\langle\theta, u\rangle|^{\alpha} m(d\theta), \alpha \in]0, 2[.$$

D.6.11 Remarks

(1) If $\alpha \neq 2$, the stable symmetric laws on $\mathbb{R}^d$ constitute a family that depends on a general symmetric measure on S^{d-1}; in the case $\alpha = 2$ they depend only on a symmetric matrix $d \times d$.

(2) Formulas (D.6.8) and (D.6.9) show that the stable PSIIs are sums of completely dissymmetric one-dimensional stable PSIIs that are carried by straight lines drawn from the origin and are independent. These sums are understood as suitable integrals.

For the exact simulation of stable laws, see Chambers et al. [1976], Devroye [1986], and Boulanger [1990].

It is also possible to simulate a stable law approximately by truncating its Lévy measure. For the processes, that means to approximate a PSII by a jump PSII. There exist also approximate simulations based on expansion by series. We quote these because of their theoretical and pratical interest; see Lepage et al. [1981] and Marcus-Pisier [1983].

Proposition D.6.12 Let X be a symmetric stable real random variable of order $\alpha, 0 < \alpha < 2$, with

$$\mathbb{E}e^{iuX} = e^{-c|u|^{\alpha}}.$$

Let $E_1, \ldots, E_n, \ldots$ be independent exponential variables with parameter 1 and let $\Gamma_n = E_1 + \cdots + E_n$.

Let V be a symmetric random variable such that

(*) $$\mathbb{E}[|V|^{\alpha}] = c\int_0^{\infty} \frac{\sin t}{t^{\alpha}}\,dt$$

and let $V_1, \ldots, V_n, \ldots$ be independent copies of V, which are independent of the E_i.

The series

$$(**) \qquad \sum_{j=1}^{\infty} (\Gamma_j)^{-(1/\alpha)} V_j$$

converges almost surely to a random variable with the same law as X.

The formula (**) is especially easy to use for simulation. The constant in (*) is given by the relation (see Feller [1966, Chapter XVII 6.4])

$$\int_0^{\infty} \frac{\sin t}{t^{\alpha}}\, dt = \Gamma(2-\alpha)\, \frac{\cos \dfrac{\pi\alpha}{2}}{\alpha - 1}, \qquad 0 < \alpha < 2.$$

The law of V can be taken to be uniform on $[-a, a]$ or discrete $\frac{1}{2}\delta_{-a} + \frac{1}{2}\delta_a$.

The law of large numbers shows that $\Gamma_j \sim j$ as $j \uparrow \infty$, and so the series (**) converges quickly when α is small.

A similar result applies to one-sided stable laws.

Proposition D.6.13 Let X be a random variable with one-sided stable law of order $\alpha \in]0, 1[$ such that

$$\mathbb{E}e^{-\lambda X} = e^{-c\lambda^{\alpha}}, \qquad \lambda > 0.$$

Let $\Gamma_1, \ldots, \Gamma_n, \ldots$ be as in the preceding proposition, and let $V_1, \ldots, V_n, \ldots$ be constructed as before from a variable V that is positive and satisfies

$$\mathbb{E}V^{\alpha} = \frac{c}{\Gamma(1-\alpha)}.$$

Then the series

$$\sum_{j=1}^{\infty} (\Gamma_j)^{-(1/\alpha)} V_j$$

converges a.s. to a random variable with the same law as X.

These representations can be extended to $\mathbb{R}^d$-valued variables and even to processes, see Marcus-Pisier [1983].

D.7 Application of the Subordination in Bochner Sense to Simulation

D.7.1 Principle of the Subordination in Bochner Sense

Let $(P_t)_{t\geq 0}$ be a Markov semi-group on the measurable space $(E, \mathscr{E})$. Let $(\mu_t)_{t\geq 0}$ be a convolution semi-group on $\mathbb{R}_+$, that is, a family of probability measures on $\mathbb{R}_+$ satisfying

$$\begin{aligned} &\mu_t * \mu_s = \mu_{t+s} && s, t \geq 0 \\ &\lim_{t\to 0} \mu_t = \delta_0 && \text{narrowly.} \end{aligned}$$

We have already seen that with such a convolution semi-group we can associate a subordinator $(Y_t)_{t\geq 0}$ that is unique in law (assuming $Y_0 = 0$), and a Bernstein function φ vanishing at zero, by the formula

$$\mathbb{E}e^{-p(Y_t - Y_0)} = \int_0^\infty e^{-py}\mu_t(dy) = e^{-t\varphi(p)}. \tag{D.7.2}$$

Let us consider the operator

$$Q_t = \int_0^\infty P_y \mu_t(dy). \tag{D.7.3}$$

Q_t is a positive operator from $b\mathscr{E}$ into itself, and $Q_t 1 = 1$. (Formula (D.7.3) is taken here in the elementary sense that for x in E and f in $b\mathscr{E}$

$$Q_t f(x) = \int_0^\infty P_y f(x)\mu_t(dy).$$

This supposes only that $t \to P_t f(x)$ be measurable.)

Moreover,

$$\begin{aligned} Q_t Q_s &= \int_0^\infty P_y \mu_t(dy) \int_0^\infty P_z \mu_s(dz) = \int P_{y+z}\mu_t(dy)\mu_s(dz) \\ &= \int_0^\infty P_x(\mu_t * \mu_s)(dx) \\ &= \int_0^\infty P_x \mu_{t+s}(dx) = Q_{t+s}. \end{aligned}$$

Thus Q_t is a Markov semi-group on $(E, \mathscr{E})$.

Let us give an intuitive explanation of how the generator of (Q_t) can be obtained from that of (P_t). If we write formally $P_t = e^{tA}$, then

$$Q_t = \int_0^\infty e^{-(-A)y} \mu_t(dy)$$

That is, by analogy with formula (D.7.2),

$$Q_t = e^{-t\varphi(-A)}. \tag{D.7.4}$$

In other words, the generator of (Q_t) is $-\varphi(-A)$. Such a formula, and the argument leading to it, can be made rigorous in some frameworks of functional analysis. This holds when (P_t) is a symmetric semi-group strongly continuous on $L^2(E, \mathscr{E}, m)$, where m is a positive σ-finite measure: see Hille-Phillips [1957] and Fukushima [1980].

D.7.5 Probabilistic Interpretation

Suppose that (P_t) is the transition semi-group of a Markov process (X_t) such that, if ρ is the initial law, we have

$$\begin{aligned} &\forall 0 < s_1 < \cdots < s_m \qquad \forall f_0, f_1, \ldots, f_m \in b\mathscr{E} \\ &\mathbb{E}[f_0(X_0) f_1(X_1) \cdots f_m(X_{s_m})] \\ &\quad = \int \rho(dx_0) f_0(x_0) \int P_{s_1}(x_0, dx_1) f_1(x_1) \\ &\quad \cdots \int P_{t_m - t_{m-1}}(x_{m-1}, dx_m) f_m(x_m). \end{aligned} \tag{D.7.6}$$

We consider a realization $(Y_t)_{t\geq 0}$ of the subordinator associated with the convolution semi-group $(\mu_t)_{t\geq 0}$ such that $Y_0 = 0$ and the process $(Y_t)_{t\geq 0}$ is independent of the process $(X_t)_{t\geq 0}$. Then we have the following result.

Proposition D.7.7 The process $Z_t = X_{Y_t}, t \geq 0$, is a realization of the semi-group Q_t: $\forall 0 < t_1 < \cdots < t_n, \forall g_0, \ldots, g_n \in b\mathscr{E}$

$$\begin{aligned} \mathbb{E}[g_0(Z_0) g_1(Z_{t_1}) \cdots g_n(Z_{t_n})] &= \int \rho(dz_0) g_0(z_0) \int Q_{t_1}(z_0, dz_1) g_1(z_1) \\ &\quad \cdots \int Q_{t_n - t_{n-1}}(z_{n-1}, dz_n) g_n(z_n). \end{aligned}$$

Proof By the independence hypothesis, we may suppose that the pair $((X_t)_{t\geq 0}, (Y_t)_{t\geq 0})$ is realized on the product of the probability spaces realizing $(X_t)_{t\geq 0}$ and realizing $(Y_t)_{t\geq 0}$. Now, by (D.7.6),

$$\mathbb{E}[g_0(Z_0)g_1(Z_{t_1})\cdots g_n(Z_{t_n})] = \mathbb{E}\left[\int \rho(dx_0)g_0(x_0)\int P_{Y_{t_1}}(x_0, dx_1)g_1(x_1) \cdots \int P_{Y_{t_n}-Y_{t_{n-1}}}(x_{n_1}, dx_n)g_n(x_n)\right].$$

Since, for every bounded measurable F,

$$\mathbb{E}[F(Y_{t_1}, Y_{t_2} - Y_{t_1}, \ldots, Y_{t_n} - Y_{t_{n-1}})] = \int F(y_1, y_2, \ldots, y_n)\mu_{t_1}(dy_1)\mu_{t_2-t_1}(dy_2)\cdots\mu_{t_n-t_{n-1}}(dy_n),$$

we obtain

$$\mathbb{E}[g_0(Z_0)g_1(Z_{t_1})\cdots g_n(Z_{t_n})] = \int \rho(dx_0)g_0(x_0)\int Q_{t_1}(x_0, dx_1)g_1(x_1) \cdots \int Q_{t_n-t_{n-1}}(x_{n_1}, dx_n)g_n(x_n). \qquad \square$$

D.7.8 Case of PSIIs

If $(P_t)_{t\geq 0}$ is the transition semi-group of a PSII $(X_t)_{t\geq 0}$ on $\mathbb{R}^d$, then it is easily seen that $Z_t = X_{Y_t}$ is a process with stationary independent increments and sample paths that are continuous on the right with left limits. Thus it is a PSII.

Let ψ be the negative definite function associated with X, given by

$$\mathbb{E}e^{i\langle u, X_t - X_0\rangle} = e^{-t\psi(u)}.$$

If we compute the function associated with Z, we get

$$\begin{aligned}\mathbb{E}e^{i\langle u, Z_t - Z_0\rangle} &= \mathbb{E}e^{i\langle u, X_{Y_t} - X_0\rangle} \\ &= \mathbb{E}e^{-Y_t\psi(u)} \\ &= e^{-t\varphi(\psi(u))},\end{aligned}$$

where φ is the Bernstein function associated with $(Y_t)_{t\geq 0}$ (which shows that Bernstein functions operate on continuous negative definite functions).

D.7.9 Example

The symmetric stable process of order $\beta = 2\alpha \in]0, 2]$ on $\mathbb{R}^d$ is obtained by subordination of the Brownian motion by the one-sided stable subordinator of order $\alpha \in]0, 1]$.

Indeed, if $\mathbb{E}e^{-pY_t} = e^{-tp^\alpha}$, $\varphi(p) = p^\alpha$, then with $\psi(u) = \frac{1}{2}|u|^2$ we get

$$\varphi \circ \psi(u) = \frac{1}{2^\alpha}|u|^{2\alpha}.$$

D.7.10 Application to the Simulation of Infinitely Divisible Laws

There are several methods for simulating infinitely divisible laws [Devroye, 1986]:

(i) the rejection method when there is an explicit density;

(ii) the rejection method when the density is only known by approximations from above and below [Devroye, 1986; Bouleau (2)];

(iii) methods related to particular forms of laws (Gaussian laws, Poisson law, etc.).

Methods (i), (ii), and (iii) allow us to simulate a set of infinitely divisible laws. We can increase this set significantly using subordination. Let us suppose that we can simulate an infinitely divisible law μ_t on $\mathbb{R}^d$, depending on the parameter t in such a way that (μ_t) is a convolution semi-group. In other words, if (X_t) is the PSII vanishing at zero associated with the semi-group (μ_t), we can simulate the random variable X_t for each fixed t. If (Y_t) is a subordinator such that we can simulate Y_1, then clearly we can simulate X_{Y_1} (by beginning with Y_1 and then simulating X_t for $t = Y_1$).

Thus, if we can simulate a marginal law of a PSII with negative definite function ψ and we can simulate the subordinator with Bernstein function φ, then one can simulate the marginal laws of the PSII with negative definite function $\varphi \circ \psi$.

Examples of $\mathbb{R}^d$-valued PSIIs are:

- A centered Brownian motion with instantaneous covariance matrix Σ. By subordinating by a one-sided stable subordinator of order α we get a symmetric stable process on $\mathbb{R}^d$ such that the measure m of formula (D.6.10) satisfies

$$\langle \Sigma u, u\rangle^\alpha = c\int_{S^{d-1}} |\langle \theta, u\rangle|^\alpha m(d\theta).$$

- A jump PSII (cf. Section D.4.5).
- A PSII whose Lévy measure is an infinite average of Dirac measures.
- An $\mathbb{R}^d$-valued PSII whose components are independent real PSII that can be simulated.
- A sum or a series of independent processes of the above types.

Examples of subordinators are:

- A subordinator that is a jump process. This includes all Bernstein functions of the form

$$\varphi_1(p) = \int_0^\infty (1 - e^{-py}) d\nu(y) \qquad \text{with} \quad \|\nu\| < +\infty.$$

We quote some forms of Bernstein functions reached in this way:

$$\varphi_{1,1}(p) = \frac{1 - e^{-pa}}{1 - \alpha e^{-pa}}, \qquad \alpha \in]0,1[, \quad a > 0$$

obtained when ν is a geometric law on $a\mathbf{N}^*$;

$$\varphi_{1,2}(p) = 1 - \exp[\theta e^{-pa} - \theta], \qquad \theta > 0, \quad a > 0$$

obtained when ν is a Poisson law;

$$\varphi_{1,3}(p) = 1 - \frac{(1-\alpha)^a e^{-pb}}{(1 - \alpha e^{-pb})^a}, \qquad a > 0, \quad \alpha \in]0,1[$$

obtained when $\nu = \sum_{n=1}^{\infty} (1-\alpha)^a a(a+1)\cdots[(a+n-2)/(n-1)!]\,\alpha^{n-1}\delta_{nb}$;

$$\varphi_{1,4}(p) = \frac{ap}{1 + bp}, \qquad a, b > 0$$

obtained when $\nu(dy) = a/b^2\, e^{-(y/b)}\, dy$;

$$\varphi_{1,5}(p) = a\left(b - \frac{1 - e^{-pb}}{p}\right), \qquad a, b > 0$$

obtained when $\nu(dy) = a 1_{[0,b]}(y)\, dy$;

$$\varphi_{1,6}(p) = a\left\{\frac{b^{n+1}}{n+1} - \frac{n!}{p^{n+1}}\left[1 - e^{-bp}\left(\frac{(bp)^n}{n!} + \frac{(bp)^{n-1}}{(n-1)!} + \cdots + 1\right)\right]\right\},$$
$$a, b > 0,$$

obtained when $\nu(dy) = 1_{[0,b]}(y).y^n dy, \qquad n \in \mathbf{N}$.

- A subordinator whose Lévy measure is an average of infinitely many Dirac masses (since then the law of Y_t is also an average of Dirac masses whose coefficients can be computed by evaluating the convolution product of the involved Poisson laws). In this way we obtain the Bernstein functions of the form

$$\varphi_2(p) = \sum_i a_i(1 - e^{-b_i p}), \qquad p \geq 0, \quad a_i, b_i \geq 0.$$

- A one-sided stable subordinator of order α. We get the Bernstein functions

$$\varphi_3(p) = ap^\alpha, \qquad a > 0, \quad \alpha \in]0, 1].$$

- A Γ-process (cf. Section D.4.3). We get the Bernstein functions

$$\varphi_4(p) = a \log(1 + bp), \qquad a, b > 0.$$

- Finally, we can take a sum of independent subordinators of the preceding types and, of course, any process obtained from the preceding subordinators by subordination. We obtain in this way the Bernstein functions of the form

$$\varphi_{i_1} \circ \varphi_{i_2} \circ \cdots \circ \varphi_{i_k},$$

where the φ_{i_j} are in the families above.

D.7.11 Application to the Simulation of PSIIs

Subordination in the Bochner sense allows us to simulate iteratively (resp., recursively) the PSII $Z_t = X_{Y_t}$ as long as we can simulate (Y_t) and we can simulate iteratively (resp., recursively) the PSII (X_t).

Let us consider the case of recursive simulation.

Suppose that $X_0 = 0$ and X_t has a density $f_t(x)$ on $\mathbb{R}^d$ (see Section C.3).

The conditional law of X_u given $X_s = a$ and $X_t = b$ is given, for $s < u < t$ and $a, b \in \mathbb{R}^d$, by

$$g_{s,u,t}(x) = \frac{f_{u-s}(x-a) f_{t-u}(b-x)}{f_{t-s}(b-a)}.$$

We use the following recursive algorithm:

Choose a point γ in the interval $[\alpha, \beta]$. Simulate Y_γ given Y_α and Y_β. The point Y_γ obtained is between Y_α and Y_β:

$$Y_\alpha < Y_\gamma < Y_\beta$$

Then simulate X_{Y_γ} given X_{Y_α} and X_{Y_β} using the formula above. This gives $g_{Y_\alpha, Y_\gamma, Y_\beta}$.

Thus, we can refine a simulated path as in classical recursive simulation.

D.8 The Bernstein Monoid

We have seen some interesting consequences for simulation of the subordination in the Bochner sense. These are mainly due to the fact that the cone of Bernstein functions is stable by composition and is therefore a monoid.

To simplify the details, we consider the cone $B_{0,1}$ of Bernstein functions vanishing at zero and taking the value 1 at 1, which we call here the Bernstein monoid. The relation

$$f \prec g \Leftrightarrow \exists h \in B_{0,1} : g = f \circ h$$

is a partial order on $B_{0,1}$. A *branch* is a maximal totally ordered subset of $B_{0,1}$. A branch that starts at the root (the identity function) and reaches φ, splits at φ into an uncountable infinity of sub-branches, which form a set isomorphic to the whole of $B_{0,1}$.

$B_{0,1}$ is therefore a self-similar or fractal object, which is particularly beautiful because of the natural place it takes in mathematics.

On the Bernstein monoid the topologies of pointwise convergence, compact convergence, and compact convergence of the functions and of their derivatives are equivalent. A branch is isomorphic under the order to a subset of $\mathbb{R}_+$. This subset can be an interval $[0, a[$ or a set with isolated points.

D.8.1 Example of Branches

(a) The stable branch. For $\alpha \in]0, 1]$ put $f_\alpha(p) = p^\alpha$, $p \in \mathbb{R}_+$. Then $\{f_\alpha, \alpha \in]0, 1]\}$ is a branch.

(b) The homographic branch. For $a \in \mathbb{R}_+$ put $h_a(p) = p(1+a)/(1+pa)$. h_a is Bernstein because $h_a = (1 + (1/a))k \circ g_a$ with $k(p) = 1 - e^{-p}$ and $g_a(p) = \log(1 + ap)$, and we have already seen that k and g_a are Bernstein.

(c) Let $\varphi_1, \ldots, \varphi_n, \ldots$ be Bernstein functions whose Lévy measures have compact supports in $\mathbb{R}_+^*$ (e.g., $\varphi_n(p) = 1 - e^{-a_n p}$) and such that $\prod_{n=1}^{\infty} \varphi_n'(1) = 0$. Then $\{\varphi_1, \varphi_1 \circ \varphi_2, \ldots, \varphi_1 \circ \varphi_2 \cdots \circ \varphi_n, \ldots\}$ is a branch [Chateau, 1990].

D.8.2 Field of All the Subordinated Processes of a Markov Process

Let $f \in B_{0,1}$. We write $(Y_x^f)_{x \geq 0}$ for the subordinator associated with the Bernstein function f such that $Y_0^f = 0$. Let us define a Markov kernel P_f from $(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+))$ into itself by putting

$$P_f u(x) = \mathbb{E}[u(Y_x^f)] \qquad \text{for} \quad u \in b\mathscr{B}(\mathbb{R}_+).$$

In particular taking $u(x) = e^{-px}$, we have

$$P_f(e^{-p\cdot})(x) = e^{-xf(p)}.$$

The kernels P_f satisfy

$$P_f P_g = P_{f \circ g} \qquad \forall f, g \in B_{0,1}.$$

By using the family $(P_f)_{f \in B_{0,1}}$, we can [Bouleau and Chateau, 1989; Chateau, 1990] construct on a single probability space a family of random variables $Y_{f,t}$ such that

- for fixed $f \in B_{0,1}$, $t \to Y_{f,t}$ is a subordinator with Bernstein function f;
- for fixed $t \in \mathbb{R}_+$, $f \to Y_{f,t}$ is a process which is Markovian along the branches with transition semi-group P_f.

More generally, if $(X_t)_{t \geq 0}$ is a Markov process and $(Y_{f,t})$ is the family chosen above, which is independent of $(X_t)_{t \geq 0}$, then the family $X_{Y_{f,t}}$ is a realization of all the subordinated processes of X, which is Markovian along the branches of $B_{0,1}$. For example, by studying the homographic

branch, it is possible to derive formulas of Pitman–Yor on some exponential martingales [Pitman and Yor, 1982]. The detailed study of the Bernstein monoid and its probabilistic interpretations is still in its early stages.

E POINT PROCESSES

In this rather applied section we build up some models involving randomly distributed objects and compute some related quantities.

E.1 Setting Out the Problem

We consider a straight girder composed of a heterogeneous material whose section S is either disc-shaped or T-shaped. Let $E(x, y)$ be a function that characterizes the material: $E(x, y) = E_1$ if (x, y) lies in a grain, $E(x, y) = E_2$ if (x, y) lies in the binder. Our aim is to compute

$$E = \frac{\int_S E(x,y)(y - y_G)^2 dx\, dy}{\int_S (y - y_G)^2 dx\, dy},$$

where $G = (x_G, y_G)$ would be the center of gravity of the section if it were homogeneous. Since the number, size, and location of the grains are random, E is random (see Figure 3.4). We want to compute $\mathbb{E}(E)$ and $\mathrm{Var}(E)$.

E.1.1 Remark

Under Saint Venant's hypotheses, it can be seen that E is equivalent Young modulus as the girder is subjected to a flexing moment with axis 0_x.

E.2 Setting Up the Models

We assume that the grains are spherical, are not cut off by the edge of the girder, and do not overlap each other. We restrict our study to the planar problem.

(a) Parametrization of the Objects

This is the first step in constructing the model: the second step is to construct a point process whose state space is the parameter space.

For the disc-shaped section, the location of a grain is defined by the

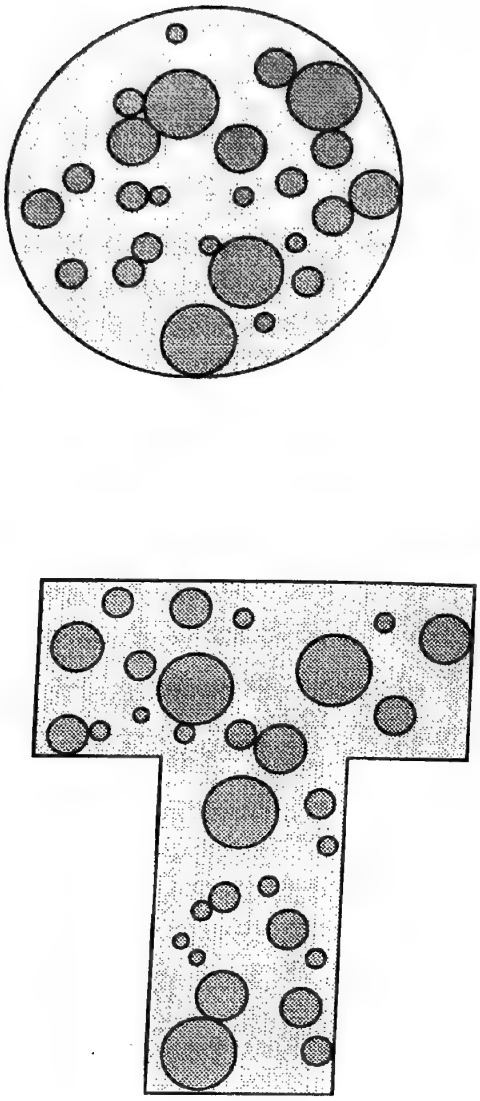


Figure 3.4

polar coordinates of its center and radius. The parameter space is

$$F = \{(\rho, \theta, a) : \rho \in [0, R], \theta \in [0, 2\pi[, a \leq R - \rho) \wedge M\},$$

where M is an upper bound for the grain radius.

For the T-shaped section, the parameter space is

$$F = \{(x, y, a) : B(x, y, a) \subset S, a \leq M\},$$

where $B(x, y, a)$ is the disc with center (x, y) and radius a.

We shall deal only with the parameter system (x, y, a).

(b) Distribution of the Point Process

Let F^* be the set of words over F:

$$F^* = \{\varnothing\} + \sum_{n=1}^{\infty} F^n.$$

Let $\mathscr{F}$ be the Borel σ-field of the metric space F. Then F^* is endowed with the σ-field

$$\mathscr{F}^* = \{A \subset F^* : A \cap F^n \in \mathscr{F}^{\otimes n}, \forall n \geq 1\}$$

Let P be an F^*-valued random variable with symmetric distribution ν, that is, for every $n \geq 2$, the restricted measure $\nu_n = \nu|_{F^n}$ is invariant under coordinate permutations. Then P determines a random distribution of points in F (of grains in our section). The probability of having n points is $\|\nu_n\| = \nu_n(F^n)$, and in that case we observe a sample of n points with distribution $\|\nu_n\|^{-1}\nu_n$.

(c) Connection with Point Processes

We denote by $M_{Pf(F)}$ the set of finite sums of Dirac measures on F. This set is endowed with the σ-field $\mathscr{M}(F)$ generated by the mappings $m \rightsquigarrow m(A), A \in \mathscr{F}$. Let $\varphi_F : F^* \to M_{Pf}(F)$ be defined by

$$\varphi_F(\{\varnothing\}) = 0, \qquad n = 0$$
$$\varphi_F(\xi_1, \ldots, \xi_n) = \sum_{i=1}^{n} \delta_{\xi_i} \qquad n \geq 1.$$

Then $N = \varphi_F \circ P$ is an $M_{Pf}(F)$-valued random variable, and this is the usual presentation of a finite point process. Intuitively, probabilities on $M_{Pf}(F)$ and symmetric probabilities on F^* are in bijective correspondance. We refer to Neveu [1977] for a rigorous proof of the following result.

Lemma E.2.1 With every probability measure σ on $(M_{Pf}(F), \mathscr{M}(F))$ we can associate a unique symmetric probability measure ν on $(F^*, \mathscr{F}^*)$ such that $\sigma = \nu \circ \varphi_F^{-1}$.

It is now clear that the distribution ν of P is completely determined by the generating functional

$$\psi_N(f) = \mathbb{E}[\exp\{-N(f)\}],$$

where $f : F \to \overline{\mathbb{R}}_+$ is a Borel function and $N(f) = \int_F f dN$. In fact we have

$$\psi_N(f) = \sum_{n=1}^{\infty} \int_{F^n} \exp\left\{ -\sum_{m=1}^{n} f(\xi_m) \right\} d\nu_n(\xi_1, \ldots, \xi_n).$$

(d) Nonoverlapping Condition
Let $\mathscr{D}_0 = \{\varnothing\}$, $\mathscr{D}_1 = F$, and, for $n \geq 2$, let $\mathscr{D}_n$ be the subset of F^n defined by

$$\mathscr{D}_n = \{((x_1, y_1, a_1), \ldots, (x_n, y_n, a_n)) : (x_i - x_j)^2 + (y_i - y_j)^2 \geq (a_i + a_j)^2, 1 \leq i < j \leq n\}.$$

The grains do not overlap iff ν_n is carried by $\mathscr{D}_n$ for all $n \geq 0$.

(e) Contact Condition
For $n \geq 2$, let $\partial\mathscr{D}_n$ be the boundary of $\mathscr{D}_n$ in the metric space F^n. If $\nu_n(\partial\mathscr{D}_n) = 0$ for every $n \geq 2$, then a.s. the grains do not touch each other. If $\nu_n(\partial\mathscr{D}_n) > 0$, several contact configurations may appear, depending on the part of the boundary that has positive measure. For instance, if $\nu_3(\partial\mathscr{D}_3) > 0$, we split up $\partial\mathscr{D}_3$ into $(\partial\mathscr{D}_3)_1, (\partial\mathscr{D}_3)_2$, and $(\partial\mathscr{D}_3)_3$, where $(\partial\mathscr{D}_3)_j$ is the subset of $\partial\mathscr{D}_3$ with exactly j equalities among the three inequalities in the definition of $\mathscr{D}_3$. If $\nu_3((\partial\mathscr{D}_3)_j) > 0$, there is a positive probability that the three grains have exactly j contact points.

E.3 Examples of Probability Distributions

(a) Poisson Point Processes
We state and comment on some classical facts about Poisson point processes, and again refer to Neveu [1977] for more details. Let $(X, \mathscr{X})$ be a measurable space such that $\{x\} \in \mathscr{X}$ for all x in X, and let μ be a finite positive measure on $(X, \mathscr{X})$.

Definition E.3.1 A *Poisson point process* with *intensity measure* μ is a measurable mapping N from some probability space $(\Omega, \mathscr{A}, \mathbb{P})$ into $(M_{Pf}(X), \mathscr{M}(X))$ such that

- for any $A \in \mathscr{X}$, the random variable $N(A)$ has Poisson distribution with mean $\mu(A)$,
- if (A_i) is a finite sequence of pairwise disjoint elements of $\mathscr{X}$, then the random variables $N(A_i)$ are independent.

We state an existence and uniqueness result.

Proposition E.3.2 For any finite positive measure μ on $(X, \mathscr{X})$ there exists a Poisson point process N with intensity measure μ. The distribution of N is uniquely determined and its generating functional is

$$\psi_N(f) = \exp\left[-\int_X (1 - e^{-f(x)})\, d\mu(x)\right].$$

Sketch of the Proof. We put

$$(\Omega, \mathscr{A}, \mathbb{P}) = (\mathbf{N}, \mathscr{P}(\mathbf{N}), m) \times (X, \mathscr{X}, \mu_0)^{\otimes \mathbf{N}},$$

where m is a Poisson distribution with mean $\theta = \|\mu\|$ and $\mu_0 = \theta^{-1}\mu$. Let $(Y, Z_1, \ldots, Z_n, \ldots)$ be the coordinate mappings on Ω. It is easily checked that the random variable

$$\begin{aligned} N(\omega) &= 0 && \text{if} \quad Y(\omega = 0 \\ &= \sum_{i=1}^{Y(\omega)} \delta_{Z_i(\omega)} && \text{if} \quad Y(\omega) \geq 1 \end{aligned}$$

has the required properties. If

$$H = \{m \in M_{Pf}(X) : m(A_1) = n_1, \ldots, m(A_k) = n_k\}$$

for $k \geq 1, A_1, \ldots, A_k$ in $\mathscr{X}$ and $n_1, \ldots, n_k$ in $\mathbf{N}$, then

$$\mathbb{P} \circ N^{-1}(H) = \mathbb{P}(N(A_1) = n_1, \ldots, N(A_k) = n_k)$$

can be computed from the hypotheses. It follows from the Caratheodory extension theorem that $\mathbb{P} \circ N^{-1}$ is uniquely determined. Regarding the generating functional, it is easily checked that for $f = \sum_{i=1}^{k} a_i 1_{A_i}, a_1, \ldots, a_k$ in $\overline{\mathbb{R}}_+, A_1, \ldots, A_k$ pairwise disjoint in $\mathscr{X}$, we have

$$\begin{aligned} \mathbb{E}[\exp -N(f)] &= \prod_{i=1}^{k} \exp\{-(1 - e^{-a_i})\mu(A_i)\} \\ &= \exp\left[-\int_X (1 - e^{-f(x)})\, d\mu(x)\right]. \end{aligned}$$

This equality extends to all Borel functions $f : X \to \overline{\mathbb{R}}_+$. □

Note that the construction of Ω above gives us a way of simulating the process N: first draw the number Y of points with Poisson distribution with mean $\theta = \|\mu\|$ and then draw independently a sample of Y points with distribution $\mu_0 = \theta^{-1}\mu$.

The framework of the proof of Proposition E.3.2 allows us also to compute the related symmetric probability measure ν of Lemma E.2.1. Put

$$\begin{aligned} P(\omega) &= \varnothing && \text{if } Y(\omega) = 0 \\ &= (Z_1(\omega), \ldots, Z_n(\omega)) && \text{if } Y(\omega) = n \geq 1. \end{aligned}$$

We have $N = \varphi_X \circ P$, and therefore

$$\begin{aligned} \|\nu_n\| = m(n) &= e^{-\theta}\,\frac{\theta^n}{n!} \\ \nu_n(A_1 \times \cdots \times A_n) &= \mathbb{P}(Y = n, Z_1 \in A_1, \ldots, Z_n \in A_n) \\ &= e^{-\theta}\,\frac{\theta^n}{n!}\,\mu_0^{\otimes n}(A_1 \times \cdots \times A_n). \end{aligned}$$

(b) Quasi-Poisson Fields of Nonoverlapping Spheres

Recall that F is a subset of $\mathbb{R}^2 \times [0, M]$. Let μ be a finite positive measure on $(F, \mathscr{F})$ and ν be the probability measure on $(F^*, \mathscr{F}^*)$ whose restriction to F^n is

$$\nu_n = \frac{c}{n!}\,1_{\mathscr{D}_n}\mu^{\otimes n}$$

(c is a normalization constant). An F^*-valued ν-distributed random variable P is called a *quasi-Poisson field* of nonoverlapping spheres.

Proposition E.3.3 Let R be a Poisson point process on F with intensity measure μ and let ρ be the symmetric probability measure on $(F^*, \mathscr{F}^*)$ such that $\rho \circ \varphi_F^{-1}$ is the distribution of R.

Put

$$\mathscr{D} = \sum_{n=0}^{\infty} \mathscr{D}_n.$$

Then

$$\nu = \frac{1}{\rho(\mathscr{D})}\,\rho|_{\mathscr{D}}.$$

Proof For any $n \in \mathbf{N}^*, A_1, \ldots, A_n$ in $\mathscr{F}$,

$$\begin{aligned}
&\rho_n|_{\mathscr{D}_n}(A_1 \times \ldots \times A_n) \\
&\quad = e^{-\theta}\frac{\theta^n}{n!}\int_{A_1\times\cdots\times A_n} 1_{\mathscr{D}_n}(\xi_1, \ldots, \xi_n)\, d\mu_0(\xi_1)\cdots d\mu_0(\xi_n) \\
&\quad = \frac{e^{-\theta}}{n!}\int_{A_1\times\cdots\times A_n} 1_{\mathscr{D}_n}(\xi_1, \ldots, \xi_n)\, d\mu(\xi_1)\cdots d\mu(\xi_n) \\
&\quad = \frac{e^{-\theta}}{c}\,\nu_n(A_1 \times \cdots \times A_n),
\end{aligned}$$

while $\rho_0(\{\varnothing\}) = e^{-\theta}, \nu_0(\{\varnothing\}) = c$. As $\|\nu\| = \|\rho\| = 1$, we get $e^{-\theta} = c\rho(\mathscr{D})$ and the result follows. □

(c) Simulation of Quasi Poisson Field

The principle is to simulate a Poisson point process and reject the outcomes whenever they do not lie in $\mathscr{D}$. In practice, we first compute the integrals

$$I_n = \int_{F^n} 1_{\mathscr{D}_n}(\xi_1, \ldots, \xi_n)\, d\mu_0(\xi_1)\cdots d\mu_0(\xi_n)$$

by sampling with distribution μ_0. This gives us the value of

$$\|\nu_n\| = \frac{I_n\theta^n}{n!}\Bigg/\sum_{m=0}^{\infty}\frac{I_m\theta^m}{m!}.$$

Numerically, we have $\|\nu_n\| = 0$ for $n > n_{\max}$, so n runs from 1 to $n_{\max}$ and given n we draw n points.

E.4 Computation of Physical Quantities

We return to the equivalent Young modulus

$$E = \frac{\int_S E(x,y)(y-y_G)^2 dx\,dy}{\int_S (y-y_G)^2 dx\,dy}$$

and the computations of $\mathbb{E}(E)$ and $\mathrm{Var}(E)$.

(a) The integrals I_n are computed by simulation. Since $n_{\max}$ is usually not too large, we can speed up the simulation for each n by using low-discrepancy sequences.

(b) Let H be a physical quantity whose value is $H_n(\xi_1, \ldots, \xi_n)$ when there are n grains with parameters $\xi_1, \ldots, \xi_n$. Then

$$\mathbb{E}(H) = \left(\sum_{p \geq 0} I_p \frac{\theta^p}{p!} \right)^{-1} \sum_{n \geq 0} \frac{\theta^n}{n!} \int_{\mathscr{D}_n} H_n(\xi_1, \ldots, \xi_n) d\mu_0(\xi_1) \cdots d\mu_0(\xi_n).$$

We must now compute

$$\int_{F^n} H_n(\xi_1, \ldots, \xi_n) 1_{\mathscr{D}_n}(\xi_1, \ldots, \xi_n) d\mu_0(\xi_1) \cdots d\mu_0(\xi_n)$$

by simulation. This can also be accelerated with low-discrepancy sequences. It is obviously necessary to use a computation procedure for $H_n(\xi_1, \ldots, \xi_n)$. If this procedure is lengthy (finite elements and so on), it becomes necessary to speed things up by using low-discrepancy sequences.

Figure 3.5 on the next page shows like Figure 3.3 an application of recursive simulation explained in Section C.2.

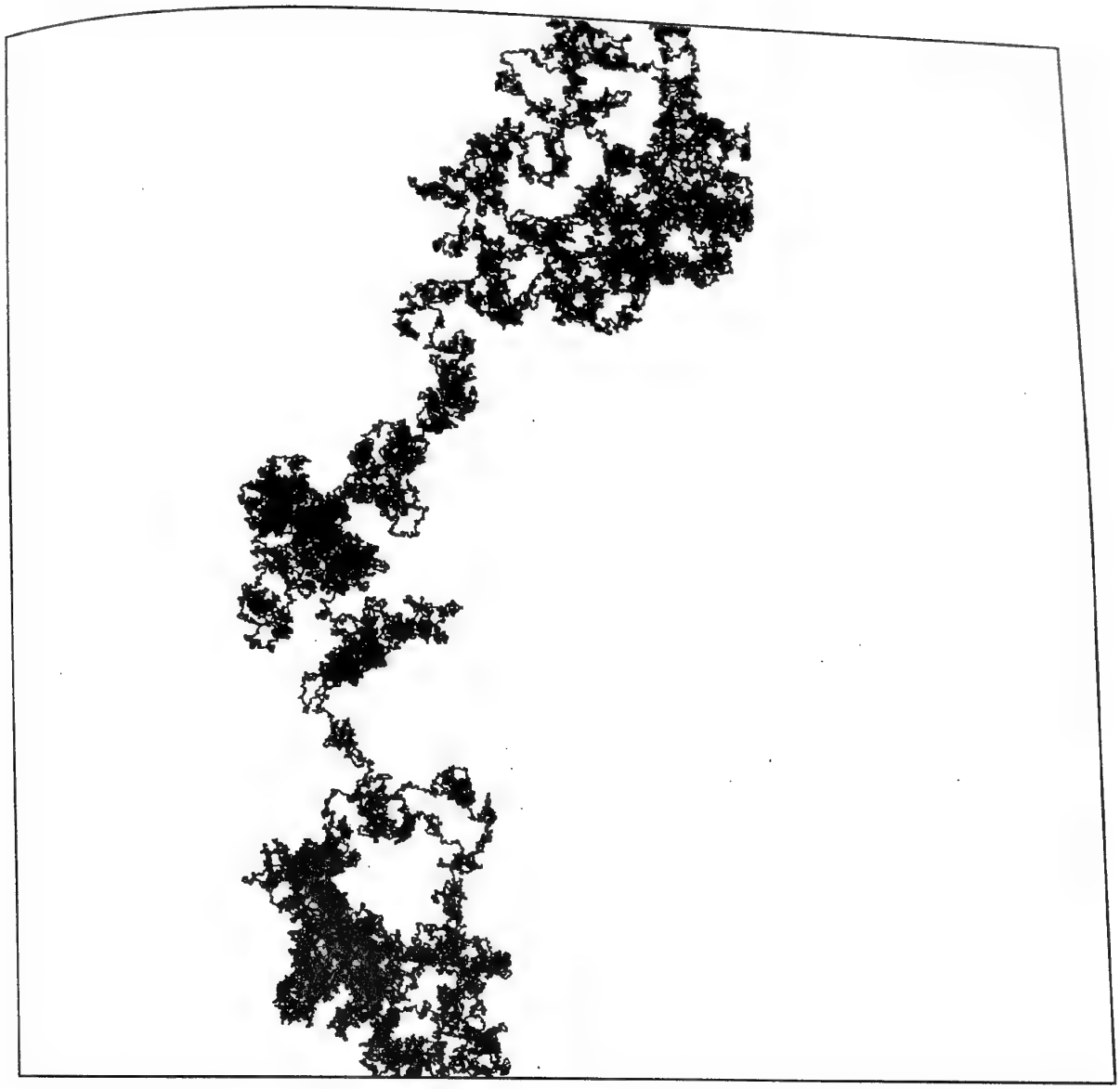


Figure 3.5 Recursive simulation of a path of a two-dimensional Brownian motion.

CHAPTER 4

Deterministic Resolution of Some Markovian Problems

In this chapter we show how methods from potential theory can be used to solve numerical problems involving Markov processes. These methods are generally faster than simulation methods; their applicability must be examined before beginning a simulation.

A ELEMENTS IN MARKOVIAN POTENTIAL THEORY

Let E be a locally compact space with a countable basis of open sets, and let $\mathscr{E}$ be its Borelian σ-field. $\mathscr{E}_+$ and $b\mathscr{E}$ denote, respectively, the cone of positive $\mathscr{E}$-measurable functions from E to $\mathbb{R}$, and the vector space of bounded $\mathscr{E}$-measurable functions from E to $\mathbb{R}$. $\mathscr{C}_0 = \mathscr{C}_0(E)$ is the Banach space of real continuous functions on E vanishing at infinity, equipped with the uniform norm $\|.\|$.

A.1 Notation and Setting: Feller Processes

We restrict ourselves to an introduction to Feller processes; for a more detailed account of the theory we refer to Meyer [1967] and, more recently, to Williams [1979] and Dellacherie and Meyer [(3)]. Other references are given in the comments.

Definition A.1.1 A *sub-Markov kernel* P on $(E, \mathscr{E})$ is a mapping from $E \times \mathscr{E}$ into $[0,1]$ such that

(i) $\forall x \in E \quad \mathscr{E} \ni A \to P(x,A)$ is σ-additive;

(ii) $\forall A \in \mathscr{E} \quad E \ni x \to P(x,A)$ is $\mathscr{E}$-measurable.

If $f \in b\mathscr{E}$, then Pf is the element of $b\mathscr{E}$ defined by

$$Pf(x) = \int f(y)P(x, dy), \qquad \forall x \in E.$$

If μ is a bounded measure on $(E, \mathscr{E})$, then μP is the bounded measure defined by

$$\mu P(A) = \int P(x, A)\mu(dx), \qquad \forall A \in \mathscr{E}.$$

Definition A.1.2 A semi-group $(P_t)_{t\geq 0}$ of sub-Markov kernels on $(E, \mathscr{E})$ is called a *Feller semi-group* if

$$\forall f \in \mathscr{C}_0(E)\, P_t f \in \mathscr{C}_0(E) \tag{A.1.3}$$

$$\forall f \in \mathscr{C}_0(E) \lim_{t\to 0} \|P_t f - f\| = 0 \tag{A.1.4}$$

Formula (A.1.3) says that P_t acts on the vector space $\mathscr{C}_0(E)$, and (A.1.4) expresses the fact that (P_t) is a strongly continuous semi-group of linear operators on the Banach space $\mathscr{C}_0(E)$.

A semi-group of sub-Markov kernels on $(E, \mathscr{E})$ satisfying (A.1.3) will also satisfy (A.1.4) provided

$$\forall f \in \mathscr{C}_0(E), \quad \forall x \in E, \quad P_t f(x) \to_{t\to 0} f(x). \tag{A.1.5}$$

This is a consequence of the important notion of resolvent family.

Lemma A.1.6 Suppose $(P_t)_{t\geq 0}$ satisfies (A.1.3) and (A.1.5). For every bounded Borel function f, the mapping

$$(t, x) \to P_t f(x) = \int f(y)P_t(x, dy)$$

is $\mathscr{B}(\mathbb{R}_+) \otimes \mathscr{E}$-measurable.

Proof This follows from the fact that, for $f \in \mathscr{C}_0, P_t f \in \mathscr{C}_0$ and $t \to P_t f(x)$ is continuous on the right. □

If a semi-group satisfies the conclusion of this lemma, it is said to be *measurable*.

Definition A.1.7 The *resolvent family* of the measurable semi-group of sub-Markov kernels (P_t) is the family of positive kernels given by

$$U_\lambda f(x) = \int_0^\infty e^{-\lambda t} P_t f(x)\, dt, \quad \lambda > 0.$$

A.1.8 Properties

- $\forall \lambda > 0, \lambda U_\lambda 1 \le 1$ and $\|\lambda U_\lambda f\| \le \|f\| \qquad \forall f \in b\mathscr{E}$.
- $U_\lambda = U_\mu + (\mu - \lambda) U_\lambda U_\mu \qquad \lambda, \mu > 0$
  This is the resolvent equation.
- The linear operators U_λ, U_μ, P_t commute.
- $\forall f \in \mathscr{E}_+, \lambda < \mu \Rightarrow U_\lambda f \ge U_\mu f$
- If there is a Borel function $f > 0$ such that $\int_0^\infty P_t f(x)\, dt < +\infty \forall x \in E$, the kernel $U = U_0 = \int_0^\infty P_t\, dt$ is said to be proper. U is the *potential kernel* of the semi-group.

Proposition A.1.9 Suppose $(P_t)_{t \ge 0}$ satisfies (A.1.3) and (A.1.5). Then $\forall \lambda > 0$, the space $U_\lambda(\mathscr{C}_0)$ is dense in $\mathscr{C}_0$, and (A.1.4) is satisfied.

Proof Clearly, $U_\lambda(\mathscr{C}_0) \subset \mathscr{C}_0$ and, by the resolvent equation, the space $U_\lambda(\mathscr{C}_0)$ does not depend on $\lambda > 0$. Now let μ be a bounded measure on E and let $f \in \mathscr{C}_0$. By (A.1.5), $\forall x \in E$

$$\lambda U_\lambda f(x) \to_{\lambda \uparrow \infty} f(x)$$

and $\lambda U_\lambda f(x)$ is bounded by $\|f\|$. Thus

$$\int \lambda U_\lambda f\, d\mu = 0 \Rightarrow \int f\, d\mu = 0$$

and, by the Hahn–Banach theorem, $U_\lambda \mathscr{C}_0$ is dense in $\mathscr{C}_0$.

But the set $\{f : \|P_t f - f\| \to_{t \to 0} 0\}$ is closed in $\mathscr{C}_0$ (Exercise). Then, from the easy fact

$$\forall f \in U_\lambda(\mathscr{C}_0) \qquad P_t f \to_{t \to 0} f \text{ in } \mathscr{C}_0$$

it follows that

$$\forall f \in \mathscr{C}_0 \qquad P_t f \to_{t \to 0} f \text{ in } \mathscr{C}_0. \qquad \square$$

A.1.10

We will use the following classical results: Let $E' = E \cup \{\delta\}$, where δ is an extra point added to E, and let $\mathscr{E}' = \mathscr{E} \cup \{B \cup \{\delta\} : B \in \mathscr{E}\}$. We define

a semi-group of Markovian kernels on $(E', \mathscr{E}')$ by

$$P_t(x, \{\delta\}) = 1 - P_t(x, E), \quad x \in \mathscr{E}$$
$$P_t(\delta, A) = 1_A(\delta), \quad A \in \mathscr{E}'.$$

This extension is still denoted by (P_t).

We write $\mathscr{E}'_+$ for the set of positive $\mathscr{E}'$-measurable functions.

Theorem A.1.11 There is a realization of the semi-group (P_t), which is continuous on the right with left limits, with life time, and strongly Markovian.

More explicitly, let $\Omega = \{\omega$: mapping from $\mathbb{R}_+$ into E' continuous on the right with left limits such that $\forall s\ \omega(s) = \delta \Rightarrow (\omega(t) = \delta \quad \forall t \geq s)\}$, let X_t be the coordinates,

$$\mathscr{F}_t^0 = \sigma(X_s, s \leq t), \qquad \text{and} \qquad \mathscr{F}_\infty^0 = \sigma(X_t, t \geq 0).$$

Then for any probability measure μ on $(E, \mathscr{E})$ there exists a probability measure $\mathbb{P}_\mu$ on $(\Omega, \mathscr{F}_\infty^0)$ such that $\forall s \leq t$

$$\text{(A.1.12)} \qquad \forall f \in \mathscr{C}_0 \quad \mathbb{E}_\mu[f(X_t)|\mathscr{F}_s^0] = P_{t-s}f(X_s) \qquad \mathbb{P}_\mu - \text{a.s.}$$

It follows by the monotone class theorem that $\forall 0 \leq t_1 < t_2 < \cdots < t_n$, for all bounded Borel functions $f_1, \ldots, f_n$ on E',

$$\mathbb{E}_\mu[f_1(X_{t_1}) \cdots f_n(X_{t_n})] = \int_{E'} \mu(dx_0) \int_{E'} P_{t_1}(x_0, dx_1) f_1(x_1) \cdots \int_{E'} P_{t_n - t_{n-1}}(x_{n-1}, dx_n) f_n(x_n).$$

The shift operators θ_t on Ω are defined by

$$\theta_t(\omega)(s) = \omega(t + s) \quad \forall \omega \in \Omega.$$

θ_t is measurable from $(\Omega, \mathscr{F}_\infty^0)$ into itself.

The property (A.1.12) can be strengthened to the Markov property: for every bounded $\mathscr{F}_\infty^0$-measurable F

(A.1.13) $$\mathbb{E}_\mu[F \circ \theta_t | \mathscr{F}_t^0] = \mathbb{E}_{X_t}[F] \qquad \mathbb{P}_\mu - \text{a.s.}$$

Notation
$\mathbb{E}_\mu, \mathbb{P}_\mu, \mathbb{E}_x, \mathbb{P}_x$ denote here the expectation (resp., probability) when the initial measure is μ (resp., δ_x, the Dirac mass at x).

Strong Markov Property. Let $\mathscr{G}_t = \mathscr{F}_{t+}^0 = \cap_{s>t} \mathscr{F}_s^0$ and let S be a $(\mathscr{G}_t)$-stopping time. Then

(A.1.14) $$\mathbb{E}_\mu[f(X_{S+t}) 1_{\{S<+\infty\}} | \mathscr{G}_S] = P_t f(X_S) 1_{\{S<+\infty\}}$$

(in the sense $\mathbb{P}_\mu$-a.s. for every μ) for every bounded Borel function f on E'. This can be strengthened to

$$\mathbb{E}_\mu[F \circ \theta_S 1_{\{S<+\infty\}} | \mathscr{G}_S] = \mathbb{E}_{X_S}[F] 1_{\{S<+\infty\}}$$

for every F bounded $\mathscr{F}_\infty^0$-measurable function F.

A.1.16 Remarks

The classical notation is as follows: $\mathscr{F}_\infty^\mu$ is the completion of $\mathscr{F}_\infty^0$ with respect to $\mathbb{P}_\mu$,

$$\mathscr{F}_\infty = \cap_\mu \mathscr{F}_\infty^\mu, \quad \mathscr{F}_t^\mu = \mathscr{F}_t^0 \vee (\mathbb{P}_\mu\text{-negligible sets of } \mathscr{F}_\infty^\mu)$$

and

$$\mathscr{F}_t = \cap_\mu \mathscr{F}_t^\mu.$$

It can be shown that the σ-fields $\mathscr{F}_t$ are continuous on the right ($\mathscr{F}_t = \mathscr{F}_{t+}$) and that $\mathscr{F}_t \supset \mathscr{F}_{t+}^0$. The strong Markov property holds under $\mathbb{P}_\mu$ for the filtration $(\mathscr{F}_t^\mu)$. These subtleties are necessary for the entrance times to be measurable: if $A \in \mathscr{E}$ then the random time

$$D_A = \inf\{t \geq 0 : X_t \in A\} \qquad (= +\infty, \text{ if } \{.\} = \varnothing)$$

is not in general an $(\mathscr{F}_{t+}^0)$-stopping time, but only an $(\mathscr{F}_t)$-stopping time.

These technicalities have no real consequence for the purpose of simulation, since every $(\mathscr{F}_t^\mu)$-stopping time is $\mathbb{P}_\mu$-almost surely equal to an $(\mathscr{F}_{t+}^0)$-stopping time (see Dellacherie and Meyer [(3), Chapter XIV, p.178]).

A.2 Fundamental Representations

A.2.1
Let f be Borel and bounded. By (A.1.12), $s \to P_{t-s}f(X_s)$ is an $(\mathscr{F}^0_s)$-martingale for $s \le t$.

A.2.2
If $f \in \mathscr{E}'_+$, then $(f(X_t))_{t\ge 0}$ is a supermartingale with respect to $(\mathscr{F}^0_{t+})$ under $\mathbb{P}_\mu \forall \mu$ if and only if $P_t f \le f \quad \forall t \ge 0$.

A.2.3
Let $f \in b\mathscr{E}'_+$. Then $e^{-\lambda s}U_\lambda f(X_s)$ is a supermartingale (for $(\mathscr{F}^0_{t+})$ under $\mathbb{P}_\mu \forall \mu$) and

$$e^{-\lambda s}U_\lambda f(X_s) = \mathbb{E}_\mu\left[\int_0^\infty e^{-\lambda t}f(X_t)\,dt \,|\, \mathscr{G}_s\right] - \int_0^s e^{-\lambda u}f(X_u)\,du.$$

A.2.4
Let T be a $(\mathscr{G}_t)$-stopping time. We put $X_\infty = \delta$ and from now on adopt the convention that if f is a function defined on E, f is extended on E' by putting $f(\delta) = 0$.

For $f \in b\mathscr{E}$ and for T $(\mathscr{G}_t)$-stopping time, let

$$P_T f(x) = \mathbb{E}_x[1_{T<+\infty} f(X_T)]$$
$$P^\lambda_T f(x) = \mathbb{E}_x[e^{-\lambda T} f(X_T)].$$

By the strong Markov property, if S, T are $(\mathscr{G}_t)$-stopping times such that $U = T + S \circ \theta_T$, then

$$\forall \lambda \ge 0 \qquad P^\lambda_U f = P^\lambda_T P^\lambda_S f.$$

A.2.5
We can extend properties A.2.3 and A.2.4 to universally measurable functions (the σ-field of universally measurable subsets of E' is the set of $A \subset E'$ such that for every probability measure μ on $(E', \mathscr{E}')$ there are $B, B' \in \mathscr{E}'$ such that $B \subset A \subset B'$ and $\mu(B' \setminus B) = 0$; this σ-field is denoted by $\mathscr{E}^*$) by replacing the filtration $(\mathscr{G}_t = \mathscr{F}^0_{t+})$ with $(\mathscr{F}_t)$.

A universally measurable function f from E to $\mathbb{R}_+$ is called λ-excessive $(\lambda \ge 0)$ if

$$\forall t \ge 0 \quad f \ge e^{-\lambda t}P_t f \quad \text{and} \quad \lim_{t\to 0} e^{-\lambda t}P_t f = f.$$

A universally measurable function f from E to $\mathbb{R}_+$ is said to be *λ-supermedian* ($\lambda \geq 0$) if

$$\forall \mu \geq 0 \quad U_{\lambda+\mu} f \leq f$$

it is said to be *semi-group-λ-supermedian* if

$$\forall t \geq 0 \quad e^{-\lambda t} P_t f \leq f.$$

(This notion is more restrictive than the preceding one, as is seen by integrating with respect to $e^{-\lambda t}\, dt$.)

The existence of a realization with sample paths continuous on the right shows, in particular, that the bounded functions that are continuous on the right on the sample paths and are semi-group-λ-supermedian are necessarily λ-excessive.

For example, the constant functions extended by zero at δ are λ-excessive, for all $\lambda \geq 0$. It can be shown [P.A. Meyer (1), Chapter IX] that the λ-excessive functions are also the λ-supermedian functions, which satisfy

$$f = \lim_{\mu \to \infty} \mu U_{\lambda+\mu} f.$$

A.2.6 Hunt's Theorem on Reduced Functions

For a Borel set A, $T_A = \inf\{t > 0, X_t \in A\}$ is an $(\mathscr{F}_t)$-stopping time. The lower envelope of λ-excessive functions ($\lambda > 0$) that dominate on A a λ-excessive function f (i.e., $\inf\{g: g\ \lambda\text{-excessive}, g \geq f \text{ on } A\}$) is equal to $P_A^\lambda f (= \mathbb{E}[e^{-\lambda T_A} f(X_{T_A})])$ except possibly for those points $z \in A$ such that $\mathbb{P}_z(T_A = 0) = 0$ (points of A that are irregular for A). If U_0 is proper, this property is valid for $\lambda = 0$.

A.2.7

The kernels

$$U_\lambda^A g(x) = \mathbb{E}_x \int_0^{T_A} e^{-\lambda t} g(X_t)\, dt,$$

for $A \in \mathscr{E}$, are the resolvent family of the process that is killed when it hits A.

A.2.8 Example

Let G be a bounded connected open set in $\mathbb{R}^d$ with Lipschitz boundary. Let $A = (\overline{G})^c$, and let (B_t) be an $\mathbb{R}^d$-valued standard Brownian motion. For

f continuous and $g \in b\mathscr{E}$, the solution of the inhomogeneous Dirichlet problem

$$\begin{cases} \lambda u - \frac{1}{2}\Delta u = g & \text{in } G \\ \lim_{x \to y} u(x) = f(y) & y \in \partial G \end{cases}$$

is given, for $\lambda \geq 0$, by

$$u(x) = \mathbb{E}_x[e^{-\lambda T_A} f(B_{T_A})] + \mathbb{E}_x\left[\int_0^{T_A} e^{-\lambda t} g(B_t)\, dt\right].$$

For a derivation, see Port and Stone, [1978].

A.2.9
An *additive functional* is a positive process (A_t) that is continuous on the right, $(\mathscr{F}_t)$-adapted, and satisfies

$$A_{t+s} = A_t + A_s \circ \theta_t.$$

It is therefore an increasing process. The function

$$u_\lambda^A(x) = \mathbb{E}_x\left[\int_0^\infty e^{-\lambda t} dA_t\right]$$

is called the λ-*potential* of (A_t). It is a λ-excessive function (exercise).

A.2.10 Example
If $h \in b\mathscr{E}_+$, $A_t = \int_0^t h(X_s)\, ds$ is an additive functional.

A.3 Feynman–Kac Formula

A.3.1 Generalized Resolvent Family

Let $h \in b\mathscr{E}_+$. We define a kernel U_h by

$$U_h f(x) = \mathbb{E}_x\left[\int_0^\infty e^{-\int_0^t h(X_s)\, ds} f(X_t)\, dt\right] \tag{A.3.2}$$

for $f \in b\mathscr{E}$. If we write I_g for the operator multiplying by $g \in b\mathscr{E}$, we have

(A.3.3) $$U_h = \sum_{n=0}^{\infty} U_\lambda (I_{\lambda - h} U_\lambda)^n \qquad \forall \lambda \geq \|h\|_\infty$$

Proof We have

$$U_h f(x) = \mathbb{E}_x \int_0^\infty e^{-\lambda t} e^{\int_0^t (\lambda - h(X_s))\, ds} f(X_t)\, dt.$$

Putting $u(X_s) = \lambda - h(X_s)$, this is equal to

$$= \mathbb{E}_x \sum_{n=0}^{\infty} \int_0^\infty \frac{1}{n!} \left(\int_0^t u(X_s)\, ds \right)^n e^{-\lambda t} f(X_t)\, dt$$

$$= \mathbb{E}_x \sum_{n=0}^{\infty} \int_{0<s_1<\cdots<s_n<t} u(X_{s_1}) \cdots$$

$$u(X_{s_n}) e^{-\lambda t} f(X_t)\, ds_1 \cdots ds_n dt$$

$$= \mathbb{E}_x \sum_{n=0}^{\infty} \int u(X_{t_1}) u(X_{t_1+t_2}) \cdots u(X_{t_1+\cdots+t_n})$$

$$e^{-\lambda(t_1+\cdots+t_n)} U_\lambda f(X_{t_1+\cdots+t_n})\, dt_1 \cdots dt_n.$$

Using the Markov property (A.1.13) gives

$$= \sum_{n=0}^{\infty} U_\lambda (I_u U_\lambda)^n f(x). \qquad \square$$

Consider a function $k \in b\mathscr{E}_+$ such that $k \geq h$. The same argument shows that

(A.3.4) $$U_h = \sum_{n=0}^{\infty} U_k (I_{k-h} U_k)^n,$$

which implies that the positive kernels U_h, U_k satisfy

(A.3.5) $$U_h = U_k + U_k I_{k-h} U_h = U_k + U_h I_{k-h} U_k.$$

This generalizes the resolvent equation.

A.3.6
Again, suppose that $h \in b\mathscr{E}_+$ and $A_t = \int_0^t h(X_s)\,ds$.
If we put

$$Q_t f(x) = \mathbb{E}_x[e^{-A_t} f(X_t)],$$

then by the additive functional property of A_t, we define a sub-Markov semi-group (Q_t) with resolvent family $(V_\lambda)_{\lambda>0}$ given by

$$V_\lambda = U_{\lambda+h}.$$

Suppose in addition that $h \in \mathscr{C}_0$. For $\alpha > \lambda + h$, (A.3.3) gives

$$V_\lambda = U_{\lambda+h} = \sum_{n=0}^{\infty} U_\lambda (I_{\alpha-\lambda-h} U_\lambda)^n, \tag{A.3.7}$$

which shows that V_λ is a continuous operator on the Banach space $\mathscr{C}_0$. The inequality $U_{\lambda+h} f \le U_\lambda f$ for $f \in b\mathscr{E}_+$, shows that $\|\lambda V_\lambda\| \le 1$.

Finally, for $f \in \mathscr{C}_0$, definition (A.3.2) gives

$$\lambda V_\lambda f(x) = \mathbb{E}_x\left[\int_0^\infty \lambda e^{-\lambda t} e^{-\int_0^t h(X_s)\,ds} f(X_t) dt\right],$$

which implies that $\lambda V_\lambda f \to f$ pointwise as $\lambda \to \infty$, and by the argument used in Proposition A.1.9, we obtain $\lambda V_\lambda f \to f$ in $\mathscr{C}_0 (V_\lambda \mathscr{C}_0 \subset \mathscr{C}_0, V_\lambda \mathscr{C}_0$ does not depend on $\lambda, V_\lambda \mathscr{C}_0$ is dense in $\mathscr{C}_0, \{f : \lambda V_\lambda f \to f$ in $\mathscr{C}_0\}$ is closed and contains $V_\mu \mathscr{C}_0$).

We quote the result:

> Every contraction resolvent family $(R_\lambda)(\|\lambda R_\lambda\| \le 1)$ on a Banach space B satisfying $\lambda R_\lambda x \to_{\lambda\uparrow\infty} x \quad \forall x \in B$ is the Laplace transform of a unique strongly continuous contraction semi-group on B.

This is a part of the Hille–Yosida theorem [Dellacherie and Meyer, 1987, Chapter VIII, Theorem 4] and shows that Q_t is a Feller semi-group. Let $(A, \mathscr{D}A)$ be the generator of (P_t), that is,

$$\mathscr{D}A = \left\{ f \in \mathscr{C}_0 : \lim_{t \to 0} \frac{P_t f - f}{t} \text{ exists in } \mathscr{C}_0 \right\}$$

and for $f \in \mathscr{D}A \quad Af = \lim_{t \to 0}(P_t f - f)/t$.

We again consider $h \in \mathscr{C}_0$, where $h \geq 0$. The generator $(B, \mathscr{D}B)$ of (Q_t) satisfies

(A.3.8) $$\mathscr{D}B = \mathscr{D}A \quad \text{and} \quad Bf = Af - hf \quad \forall f \in \mathscr{D}A.$$

Indeed, it is easy to see that $\mathscr{D}A = U_\lambda \mathscr{C}_0$, which does not depend on $\lambda > 0$, and by formula (A.3.4) $V_\lambda \mathscr{C}_0 = U_\lambda \mathscr{C}_0$. But if $f \in \mathscr{D}A$, then we have

$$\frac{1}{t}(Q_t f(x) - f(x)) = \mathbb{E}_x \left[f(X_t) \frac{e^{-A_t} - 1}{t} \right] + \frac{1}{t} \mathbb{E}_x [f(X_t) - f(x)],$$

which converges pointwise to

$$-f(x)h(x) + Af(x).$$

Thus the limit in $\mathscr{C}_0$ exists and is equal to $Af - hf$.

To sum up, we have the following result, which is known as the Feynman–Kac formula:

Let $h \in \mathscr{C}_0$ and let A_t be the additive functional $\int_0^t h(X_s)\, ds$ of the Feller process $(X_t)_{t \geq 0}$ with semi-group (P_t) and generator $(A, \mathcal{D}A)$.

Let $(Q_t)_{t \geq 0}$ be the semi-group defined by

$$Q_t f(x) = \mathbb{E}_x [e^{-A_t} f(X_t)] \qquad f \in \mathscr{C}_0.$$

Then (Q_t) is a Feller semi-group with generator $(B, \mathscr{D}B)$ given by

$$\mathscr{D}B = \mathscr{D}A \quad \text{and} \quad Bf = Af - hf \quad \forall f \in \mathscr{D}A.$$

In its various guises the F-K formula is ubiquitous throughout much of quantum physics on the one hand and probability theory on the other. ...

I find Feynman's formula to be beautiful. It connects the quantum mechanical propagator, which is a twentieth-century concept, with the classical mechanics of Newton and Lagrange in a unique compelling way.

My own formula is more prosaic but it, too, spans two centuries: the differential equation my formula solves is nineteenth-century, while the Wiener integral is a creation of the twentieth. [Mark Kac, 1985]

A.3.9 Example

Let G be a bounded connected open set in $\mathbb{R}^d$ with Lipschitz boundary and let $A = (\overline{G})^c$. Let f, g, h be in $\mathscr{C}_0(\mathbb{R}^d), h \geq 0$. The solution of the problem

$$\begin{cases} h.u - \frac{1}{2}\Delta u = g \text{ in } G \\ \lim\limits_{x \to y} u(x) = f(y) \qquad y \in \partial G \end{cases}$$

is given by

$$u(x) = \mathbb{E}_x[e^{-\int_0^{T_A} h(B_s)\,ds} f(X_{T_A})] + \mathbb{E}_x\left[\int_0^{T_A} e^{-\int_0^t h(B_s)\,ds} g(B_t)\,dt\right].$$

(For an extension of this example see Friedmann [1976] and Port and Stone [1978].)

Under suitable assumptions, we can extend to h Borel or even replace h by a measure [Feyel and Pradelle, 1988]. We can also take h to be complex, which enables us to interpret the Feynman integrals in quantum physics. See, for example, Hu-Meyer [(1)] and the references in it.

B BALAYAGE ALGORITHMS

We tackle only the elementary theory of balayage (or sweeping), which concerns functions. The general theory deals with balayage of measures or of the corresponding additive functionals [Dellacherie and Meyer, 1985, Blietner-Hansen, 1986].

B.1 Poincaré's Balayage

Henri Poincaré discovered that knowing the solution of the Dirichlet problem for spheres and knowing the uniqueness of that solution allow us to solve the problem for open sets of general shape, provided that they have sufficiently smooth boundary. This idea lies behind the local axiomatic potential theories, [Brelot, 1966; Bauer, 1966; Smyrnelis, 1976]. Our main interest is the algorithmic nature of Poincaré's method. For a

contemporary presentation of the Newtonian potential theory, see Brelot [1959].

(a) Poisson's Integral

We can prove the following results with arguments similar to those used for the half-space in Chapter 3, Section C.4: Let f be a continuous real function on the sphere $S(0,R)$, that is the boundary of the open ball $B(0,R)$ of center 0 and radius R in $\mathbb{R}^n, n \geq 2$. Then the function $H^S(f)$, which is harmonic in $B(0,R)$ and converges to f at the boundary, is given by

$$H^S(f)(x) = \frac{1}{\alpha_n R} \int_S \frac{R^2 - |x|^2}{|y-x|^n} f(y)\, d\sigma(y),$$

where σ is the area measure on S, and α_n is the area of $S(0,1)$ given by $\alpha_n = n(\pi^{n/2}/\Gamma((n/2)+1))$; that is,

$$\alpha_{2p} = 2\frac{\pi^p}{(p-1)!}, \quad \alpha_{2p+1} = 2\frac{(2\pi)^p}{1.3.5\ldots(2p-1)} \qquad p \geq 1.$$

Similarly, the function $K^S(f)$ defined outside $B(0,R)$ by

$$K^S(f)(x) = \frac{1}{\alpha_n R} \int_S \frac{|x|^2 - R^2}{|y-x|^n} f(y) d\sigma(y)$$

is harmonic in the open set $(\overline{B(0,R)})^c$ and converges to f at the boundary. We have

$$\|H^S(f)\|_\infty \leq \|f\|_\infty, \quad \|K^S(f)\|_\infty \leq \|f\|_\infty$$

and

$$\lim_{x\to\infty} K^S(f)(x) = 0 \qquad \text{if} \quad n \geq 3$$
$$= \int_S f(y)\frac{d\sigma(y)}{2\pi R} \quad \text{if} \quad n = 2.$$

Definition B.1.1 A function f is said to be *continuous superharmonic* in an open set $G \subset \mathbb{R}^n$ if it is continuous in G and, for every closed ball contained in G with center x and boundary S, it satisfies $f(x) \geq 1/\sigma(S) \int_S f(y) d\sigma(y)$.

Lemma B.1.2 Let f be continuous in the open set G.

(i) If, for every $x \in G$, there exists $R(x) > 0$ s.t.

$$f(x) \geq \frac{1}{\sigma(S)} \int_S f(y)\, d\sigma(y)$$

for every sphere S that is the boundary of a closed ball contained in G with center x and radius $r < R(x)$, then f is continuous superharmonic in G.

(ii) Let $\overline{B}$ be a closed ball contained in G with boundary S. The function $\tilde{f}$, obtained by replacing f with its Poisson integral $H^S(f)$, is continuous superharmonic whenever f is superharmonic.

We refer to Brelot [1959] for the proof.

The function $\tilde{f}$ is called the *balayée* of f outside B. (The word comes from the fact that it can be shown that the positive measure $-\Delta\tilde{f}$ is obtained from the positive measure $-\Delta f$ by suitably distributing the masses from B onto the boundary S of B, without changing masses outside B or the total mass.)

(b) Balayage in an Open Set Satisfying the Outside Sphere Condition

Let Ω be a bounded connected open set in $\mathbb{R}^n$. Consider a continuous function f on the boundary $\partial\Omega$ of Ω. We are looking for a function $H(f)$ that is harmonic in Ω and tends to f at the boundary. The uniqueness follows from the fact that if a harmonic function in a connected open set reaches its maximum inside the open set, then it is constant on the open set.

The main problem is therefore existence. It is sufficient, for this purpose, to consider the case when f is the trace on $\partial\Omega$ of a positive function F which is C^2 on $\mathbb{R}^n$ and satisfies $\Delta F < 0$ in a ball A containing $\overline{\Omega}$. Indeed, the continuous function f on the compact set $\partial\Omega$ is a uniform limit of polynomials p_n. If the functions $H(p_n)$ exist, then by the properties of convergence of harmonic functions [Brelot, 1959], the functions $H(p_n)$ converge uniformly in $\overline{\Omega}$ to a harmonic function, which must be $H(f)$. But every polynomial is the difference of two polynomials p such that $p > 0$ and $\Delta p < 0$ in A.

Now consider a covering of Ω by open balls $B_1, B_2, \ldots, B_k, \ldots$ with boundaries contained in Ω. We enumerate these balls in any order, provided that every ball occurs infinitely many times, for example, in the order $B_1, B_2, B_1, B_2, B_3, B_1, B_2, B_3, B_4, \ldots$. Let $B^1, B^2, B^3, B^4, \ldots, B^n, \ldots$ be

this enumeration. If we consider the successive balayées $F_1, F_2, \ldots, F_n, \ldots$ of F outside these balls, where F_n is the balayée of F_{n-1} outside B^n, we obtain a decreasing sequence $F, F_1 \ldots, F_n, \ldots$, of positive continuous functions in B that are superharmonic in Ω and equal F on $A\backslash\Omega$. Let F_∞ be the limit of the F_n. In every open ball B such that $\overline{B} \subset \Omega, F_\infty$ is the decreasing limit of positive harmonic functions and so is harmonic. Thus F_∞ is harmonic in Ω.

Let us suppose in addition that it is possible to put a sphere S outside Ω that is contained in A and contains the point $y \in \partial\Omega$. Solving the outer Dirichlet problem for S with F as boundary value gives $K^S(F)$. Let U equal $K^S(F)$ outside S, and F on S and inside S. Then U is harmonic outside S and hence in Ω, and is continuous in A. Finally, it follows easily that

$$U \leq F_n \leq F \qquad \forall n,$$

and hence

$$U \leq F_\infty \leq F$$

F_∞ is sandwiched between two continuous functions that coincide at y and is therefore continuous at y.

Thus if the open set Ω satisfies the outer sphere condition at every point of its boundary, then F_∞ solves the Dirichlet problem for F.

B.2 The Harmonic Kernel of a Markov Chain

Let $(E, \mathscr{E})$ be a measurable space and let P be a sub-Markov kernel on $(E, \mathscr{E})$. The potential kernel of P is defined by

$$G = \sum_{n=0}^{\infty} P^n \qquad (P^0 = I).$$

For $A \in \mathscr{E}$, let I_A denote the operator multiplying by the indicator function 1_A. We define

$$G_A = \sum_{n=0}^{\infty} (PI_A)^n. \tag{B.2.1}$$

Introducing, as usual, an extra point δ, let (X_n) be a realization of a Markov chain with values in $E \cup \{\delta\}$ and transition P. (We adopt the convention that $u(\delta) = 0$ for $u \in b\mathscr{E}$ and $X_\infty = \delta$.) We introduce the kernels

(B.2.2) $$H_A u(x) = \mathbb{E}_x[u(X_{T_A})], \qquad u \in b\mathscr{E},$$

where $T_A = \inf\{n \geq 0, X_n \in A\}$ is the hitting time of A.
We get

$$H_A u(x) = \mathbb{E}_x\left[\sum_{n=0}^{\infty} u(X_n) 1_{X_0 \notin A} \cdots 1_{X_{n-1} \notin A} 1_{X_n \in A}\right]$$
$$= u(x)1_A(x) + \sum I_{A^c} P I_{A^c} \cdots I_{A^c} P(1_A u)(x)$$

and so

(B.2.3) $$H_A = I_A + I_{A^c} G_{A^c} P I_A.$$

Let us state some properties of the kernels H_A: We say that u is *excessive* for P if u is $\mathscr{E}$-measurable from E to $\overline{\mathbb{R}}_+$ and $u \geq Pu$.

- $\forall k \in \mathbb{N}, v = u \wedge k$ is excessive and finite, and so $v(X_n)$ is a positive supermartingale and $H_A v(x) = \mathbb{E}_x v(X_{T_A}) \leq v(x) = \mathbb{E}_x[v(X_0)]$. Letting $k \uparrow \infty$, we obtain

  (B.2.4) $$H_A u \leq u \qquad \forall u \text{ excessive.}$$

- Let θ denote the translation operator of the Markov chain (X_n). We have $T_A = 1 + T_A \circ \theta$ on $\{T_A > 0\}$; hence by the Markov property,

$$H_A u(x) = 1_A.u(x) + \mathbb{E}_x[u(X_{T_A})1_{T_A>0}]$$
$$= 1_A.u(x) + \mathbb{E}_x[\mathbb{E}_{X_1}[u(X_{T_A})]1_{T_A>0}]$$
$$= 1_A.u(x) + I_{A^c} P H_A u(x)$$

  and so

  (B.2.5) $$H_A = I_A + I_{A^c} P H_A.$$

- (B.2.6) If u is excessive, then $v = H_A u$ is excessive. Indeed, by (B.2.4), $Pv \leq Pu \leq u$, which equals to v on A. By (B.2.5) we have $v = Pv$ on A^c, and thus $v \leq Pv$ everywhere.
- (B.2.7) If w is excessive and $w \geq u$ on A, then $H_A u \leq H_A w$, since measures $H_A(x, .)$ are carried by A, and so $H_A u \leq w$ everywhere.
- Applying P on the left in (B.2.3) gives

$$PH_A = PI_A + PI_{A^c}G_{A^c}PI_A = G_{A^c}PI_A$$

and hence

(B.2.8) $$PH_A = G_{A^c}PI_A.$$

Now the general identity

$$G(P_1 + P_2) = G(P_1) + G(P_1)P_2G(P_1 + P_2)$$

between potential kernels [Dellacherie and Meyer (2)], applied with $P_1 = I_{A^c}P$ and $P_2 = I_AP$, gives

(B.2.9) $$G = H_AG = I_{A^c}G_{A^c}.$$

We have thus proved the following theorem:

Theorem B.2.10 There exists a sub-Markov kernel H_A such that for every excessive function u, H_Au is the smallest excessive function larger than u on A and

(i) $H_A = I_A + I_{A^c}G_{A^c}PI_A$
(ii) $H_A = I_A + I_{A^c}PH_A$
(iii) $PH_A = G_{A^c}PI_A$
(iv) $G = H_AG + I_{A^c}G_{A^c}$.

Corollary B.2.11 (Domination Principle) Let v be excessive and let $f \geq 0$ be measurable. If $v \geq Gf$ on $\{f > 0\}$ then $v \geq Gf$ everywhere.

Proof Let $A = \{f > 0\}$. From $v \geq Gf$ on A we get $H_Av \geq H_AGf$. But by part (iv) of the theorem, $H_AGf = Gf$ and hence

$$v \geq H_Av \geq H_AGf = Gf.$$ □

B.3 Numerical Balayage

A classical problem in geodesy is the leveling problem.

E is a finite set that can be assumed to be contained in $\mathbb{R}^2$. For every $x \in E$, differences of altitude $h_x - h_y = \alpha_{xy}$, $\alpha_{xy} = -\alpha_{xy}$ have been measured between x and some other points $y \in E$ (neighbor of x or not). But these measurements are not compatible because of the errors; that is, there exists

no function h from E to $\mathbb{R}$ such that

$$h(x) - h(y) = \alpha_{xy} \qquad \forall x, y.$$

If we choose a quadratic criterion, then we must find an h which minimizes the sum

$$\sum_{(x,y)\in M} (h(x) - h(y) - \alpha_{xy})^2,$$

where the symmetric relation $M \subset E^2$ denotes the pairs (x, y) for which a measurement has been taken. The sum can also be taken with weights

$$\sum_{(x,y)\in M} q(x, y)(h(x) - h(y) - \alpha_{xy})^2. \tag{B.3.1}$$

We can make arguments for such a quadratic criterion by using likelihood ratios and assuming that the errors are small and Gaussian [Poincaré, 1912], but this is not a proof, as the celebrated aphorism about the least-squares method says:

> Tout le monde y croit parce que les mathématiciens s'imaginent que c'est un fait d'observation et que les observateurs que c'est un théorème de mathématiques.[1]
> [Henri Poincaré, 1912]

We take the form (B.3.1) as a starting point. Putting $M_x = \{y: (x, y) \in M\}$, we have

$$\sum_{y\in M_x} q(x, y)(h(x) - h(y) - \alpha_{xy}) = 0 \qquad \forall x \in E. \tag{B.3.2}$$

If we now consider the stochastic matrix

$$P(x, y) = \frac{q(x, y)}{\sum_{y\in M_x} q(x, y)}, \qquad P(x, y) = 0 \quad \text{if} \quad y \notin M_x, \tag{B.3.3}$$

[1]Everyone believes it, because mathematicians think that it is a fact of observation and observers think that it is a theorem of mathematics.

then equation (B.3.2) becomes

$$(P - I)h = -\beta,$$

where

$$\beta(x) = \left(\sum_{y \in M_x} q(x,y)\alpha_{xy} \right) \Big/ \sum_{y \in M_x} q(x,y).$$

Clearly, to construct a relief map, we must also use boundary conditions. So we suppose that the values of h are given at some points in E that we call the boundary and denote by ∂E. The problem is then to find h satisfying

$$\text{(B.3.4)} \qquad \begin{cases} (P - I)h = -\beta & \text{in } E \backslash \partial E \\ h = f & \text{in } \partial E. \end{cases}$$

The matrix $P(x,y)$ in this problem is used only for points in $E\backslash\partial E$, so for convenience we change the definition of P, keeping (B.3.3) for $x \in E\backslash\partial E$ and putting

$$\begin{aligned} P(x,x) &= 1 \\ P(x,y) &= 0 \quad y \neq x \end{aligned}$$

for $x \in \partial E$.

B.3.5 The Homogeneous Problem

We consider first the homogeneous problem

$$\begin{cases} (P - I)h = 0 & \text{in } E \backslash \partial E \\ h = f & \text{in } \partial E. \end{cases}$$

Let η_0 be the function defined by

$$\eta_0 = \|f\|_\infty 1_{E\backslash\partial E} + f 1_{\partial E};$$

this clearly satisfies $P\eta_0 \leq \eta_0$.

We define functions η_n recursively by

$$\eta_{n+1} = K_n \eta_n,$$

where the operator K_n is defined by

$$K_n g = g 1_{E\setminus\{x_n\}} + (Pg(x_n))1_{\{x_n\}}$$

and (x_n) is a sequence of points of $E\setminus\partial E$ that contains every point in $E\setminus\partial E$ infinitely often.

Lemma B.3.6 If ξ is a function from E into $\mathbb{R}$ such that

$$P\xi \le \xi,$$

then $PK_n\xi \le K_n\xi$.

Proof (a) From $\xi \ge P\xi$ follows that, outside $\{x_n\}$,

$$K_n\xi \ge P\xi \ge PK_n\xi.$$

(b) At x_n,

$$K_n\xi(x_n) = P\xi(x_n) \ge PK_n\xi(x_n).$$

The sequence η_n is decreasing by the lemma and is bounded below by $-\|f\|_\infty$. Let η_∞ be its limit. Since $\eta_n \ge P\eta_n$, we have $\eta_\infty \ge P\eta_\infty$. Now at x_n we have $\eta_{n+1}(x_n) = P\eta_n(x_n)$, and so for any fixed point x in $E\setminus\partial E$ there are infinitely many indices n such that $\eta_{n+1}(x) = P\eta_n(x)$. Thus η_∞ is also the limit of the subsequence of the η_n satisfying

$$\eta_{n+1}(x) = P\eta_n(x),$$

and η_∞ satisfies

$$\eta_\infty(x) = P\eta_\infty(x) \qquad \forall x \in E\setminus\partial E.$$

But $\eta_\infty(x) = f(x)\ \forall x \in \partial E$, since $\eta_n(x) = f(x)\ \forall x \in \partial E, \forall n$. Therefore η_∞ is a solution of (B.3.4). □

With the assumptions above, this solution may not be unique. Uniqueness can be proved if $\forall x \in E\setminus\partial E$ the probability that the Markov chain with transition P starting at x hits the boundary ∂E is strictly positive.

B.3.7 The Inhomogeneous Problem

We now turn to the problem

$$\text{(B.3.8)} \qquad \begin{cases} (P-I)h = -\beta & \text{in } E\backslash\partial E \\ h = 0 & \text{in } \partial E. \end{cases}$$

Putting $\beta = \beta^+ - \beta^-$, we may assume that $\beta \geq 0$. Suppose we have found a function ξ_0 such that

$$\xi_0 \geq P\xi_0 + \beta$$

and

$$\xi_0 = 0 \qquad \text{on} \quad \partial E.$$

The algorithm

$$\xi_{n+1} = J_n \xi_n,$$

with J_n defined by

$$J_n g = g 1_{E\backslash\{x_n\}} + (Pg(x_n) + \beta(x_n)) 1_{\{x_n\}}$$

for a sequence (x_n) of points in $E\backslash\partial E$ that contains every point in $E\backslash\partial E$ infinitely often, converges to a solution ξ_∞ of the problem (B.3.8).

The argument is similar to the preceding one. A sufficient condition for (B.3.8) to have a unique solution is that there exists $\psi > 0$ satisfying $\psi - P\psi > 0$ in $E\backslash\partial E$.

B.3.9 Comment

Such balayage algorithms are easily implemented and the property of monotony makes them especially efficient. They can be generalized to some nonlinear operators P.

C REDUCED FUNCTION ALGORITHM

C.1 Reduced Function Algorithm with Respect to a Sub-Markov Kernel

Let $(E, \mathscr{E})$ be a measurable space and let P be a sub-Markov kernel on $(E, \mathscr{E})$. A measurable positive function f is called excessive for P if $f \geq Pf$.

Definition C.1.1 Let u be a bounded measurable function. The function Ru defined by

$$Ru = \inf\{f \text{ excessive}, f \geq u\}$$

is called the *reduced function* or *réduite* of u.

Proposition C.1.2 Let u be measurable and bounded. If $u_0 = u^+$ and $u_{n+1} = u_n \vee Pu_n$, then the sequence u_n is increasing and converges to Ru.

Proof Let $w = \sup_n u_n$. We have

$$Pw = \sup_n Pu_n \leq \sup_n u_{n+1} = w,$$

and so w is excessive.

Now, if v is excessive and $v \geq u$, we have $v \geq u^+$ and $v \geq u_n \Rightarrow Pv \geq Pu_n$. Hence $v \geq u_n \vee Pu_n = u_{n+1}$, and so $v \geq w$. □

Corollary C.1.3

(i) We also have $u_{n+1} = u \vee Pu_n$.

(ii) Ru is measurable.

(iii) $Ru = u \vee PRu$.

(iv) $Ru = PRu$ on the set $\{u < Ru\} \cup \{u \leq 0\}$.

Proof (i) Let $v_n = u \vee Pu_{n-1}$. Then (v_n) increases with (u_n), and if $u_n = v_n$ then

$$v_{n+1} = u \vee Pu_n \leq v_n \vee Pu_n = u_{n+1}$$
$$u_{n+1} = v_n \vee Pv_n \leq v_{n+1} \vee Pv_n = v_{n+1}.$$

Part (ii) comes from Proposition C.1.2, and part (iii) follows from part (i). □
(iv) By (iii), we have $Ru = PRu$ on $\{u < Ru\} \cup \{u \leq 0\}$.

C.1.4

We quote some simple properties of the (nonlinear) réduite operator $R: u \to Ru$.

- R is increasing ($u \leq v \Rightarrow Ru \leq Rv$), subadditive ($R(u+v) \leq Ru + Rv$) and positively homogeneous ($R(tu) = tRu, t \in \mathbb{R}_+$).
- If $u_n \uparrow u$ then $Ru_n \uparrow Ru$, by Definition C.1.1.

- If $|u - v| \le h$ for some excessive function h, we have $u \le v + h$, and hence $Ru \le Rv + h$; similarly, $Rv \le Ru + h$. Thus $|Ru - Rv| \le h$. In particular, this is true when h is constant.

C.2 The Filling Scheme and the Réduite Property

Let us consider the action of the réduite algorithm (Proposition C.1.2) on a function u of the form $u = Gf$, where $G|f| < +\infty$. We denote by K the nonlinear operator $v \to v \vee Pv$. As usual, we let $f^+ = f1_{\{f \ge 0\}}$, and $f^- = -f1_{\{f \le 0\}}$. Then

$$\begin{aligned}(I - P)KGf &= (I - P)[Gf \vee (Gf - f)] = (I - P)(Gf + f^-) \\ &= f^+ - P(f^-).\end{aligned}$$

Hence the operator $S = (I - P)KG$ acts on functions f such that $G|f| < +\infty$ by sweeping by P the negative part of f. S is called the *filling operator*.

Clearly the sequence $(S^n f)^+$ is decreasing and the sequence $(S^n f)^-$ satisfies

$$(S^n f)^- \le P^n(f^-)$$

and so, tends to zero since $G|f| < +\infty$.

If we put $h = \lim_n S^n f = \lim_n (S^n f)^+$, then $0 \le h \le f^+$ and

$$Gh = \lim_n GS^n f = \lim_n K^n Gf = Ru.$$

We have proved the following proposition.

Proposition C.2.1 Let u be a bounded measurable function of the form $u = Gf$, where f satisfies $G|f| < +\infty$. Then $Ru = Gh$ for some h such that $0 \le h \le f^+$.

Corollary C.2.2 If u and v are bounded and excessive then $u - R(u - v)$ is excessive.

Proof (a) Suppose first that $P1 \le \alpha < 1$. Then $w = G(I - P)w$ for bounded w, and thus $u = Gf$ and $v = Gg$ with $f = (I - P)u$ and $g = (I - P)v$, which are positive. By Proposition C.2.1, $R(u - v) = Gh$, with $0 \le h \le (f - g)^+$, and hence $h \le f$ and $u - R(u - v) = G(f - h)$ is excessive.

(b) We assume only that P is sub-Markov, and write ${}^\alpha P = \alpha P$ for $\alpha \in]0, 1[$. Then $u - {}^\alpha R(u - v)$ is ${}^\alpha P$-excessive by part (a). When $\alpha \uparrow 1$, it is easy to see that ${}^\alpha Rw \uparrow Rw$ for any bounded w. Since $u - {}^\alpha R(u - v)$ is

${}^{\beta}P$-excessive for $\beta < \alpha$, the limit $u - R(u - v)$ is still ${}^{\beta}P$-excessive. This is true for any $\beta \in]0, 1[$, so $u - R(u - v)$ is P-excessive. □

The réduite property stated in Corollary C.2.2 was shown by G. Mokobodzki to be characteristic of the cones (such as the cone of excessive functions with respect to a sub-Markov kernel P) from which a potential theory can be constructed [Mokobodzki and Sibony, 1968; Mokobodzki, 1970].

C.3 Optimal Stopping

Recall first the idea of an essential sup of an eventually uncountable family of random variables.

Definition-Proposition C.3.1 Let $(Y_i)_{i\in I}$ a family of measurable functions from (Ω, F) into $(\overline{\mathbb{R}}_+, \mathscr{B}(\overline{\mathbb{R}}_+))$. There exists a random variable Z unique up to $\mathbb{P}$-a.s. equality such that

(i) $Z \geq Y_i$ a.s. $\forall i \in I$;

(ii) for any random variable H such that $H \geq Y_i$ a.s. $\forall i \in I$, we have $H \geq Z$ a.s.

The random variable Z is denoted ess $\sup_{i\in I} Y_i$. There exists a sequence in I such that $Y_{i_n} \to Z$ $\mathbb{P}$-a.s. The sequence $(Y_{i_n})_{n\geq 0}$ can be chosen increasing if the family $(Y_i)_{i\in I}$ has the following stability property: $\forall i, j \in I, \exists k \in I$ s.t. $Y_k \geq Y_i \vee Y_j$ a.s.

See Neveu [1972] for a proof.

The idea of réduite possesses a probabilistic interpretation in terms of optimal stopping. We give only an introduction to this large area. For a deeper study we refer to El Karoui [1981].

Let (X_n) be a Markov chain with values in $(E, \mathscr{E})$ and transition probability P. Let u be a bounded positive measurable function such that the present benefit of some decision at time n is represented by $u(X_n)$. We are interested in evaluating

$$\sup_{T\in\mathscr{T}} \mathbb{E}_x[u(X_T)],$$

where $\mathscr{T}$ is the set of bounded stopping times for the filtration $\mathscr{F}_n = \sigma(X_m, m \leq n)$.

Proposition C.3.2

$$\sup_{T \in \mathcal{T}} \mathbb{E}_x[u(X_T)] = Ru(x).$$

Proof (a) Ru is excessive and $Ru \geq u$, so the sequence $Ru(X_n)$ is a super-martingale and $Ru(X_n) \geq u(X_n)$ for every $n \geq 0$.

(b) Let Y be a supermartingale such that $Y_n \geq u(X_n) \forall n \geq 0$. Using the notation of C.1.2,

$$\left.\begin{array}{l} Y_n \geq \mathbb{E}[Y_{n+1}|\mathcal{F}_n] \\ Y_n \geq u_k(X_n) \quad \forall n \end{array}\right\} \Rightarrow \begin{array}{l} Y_n \geq \mathbb{E}[u_k(X_{n+1})|\mathcal{F}_n] = Pu_k(X_n) \\ \text{hence } Y_n \geq u_{k+1}(X_n) \end{array}$$

Therefore, by induction on k,

$$Y_n \geq Ru(X_n) \quad \forall n \geq 0.$$

(c) Since $u \leq Ru$, we have

$$\mathbb{E}_x[u(X_T)] \leq \mathbb{E}_x[Ru(X_T)] \leq Ru(x)$$

because $Ru(X_n)$ is a positive supermartingale. Hence

$$\sup_{T \in \mathcal{T}} \mathbb{E}_x[u(X_T)] \leq Ru(x).$$

(d) For the converse inequality, we put

$$\mathcal{T}_n = \{T \in \mathcal{T}, T \geq n\}$$
$$V_n = \operatorname{ess\,sup}_{T \in \mathcal{T}_n} \mathbb{E}[u(X_T)|\mathcal{F}_n].$$

We check first that the family of random variables

$$\{\mathbb{E}[u(X_T)|\mathcal{F}_n], T \in \mathcal{T}_n\}$$

is stable by taking the sup: (consider $T = 1_{A^c}.T_1 + 1_A.T_2$ with

$$A = \{\mathbb{E}[u(X_{T_1})|\mathcal{F}_n] < \mathbb{E}[u(X_{T_2})|\mathcal{F}_n]\}).$$

This shows the existence of a sequence $(T_k^n)_{k\geq 0}, T_k^n \in \mathscr{T}_n$, such that

$$\mathbb{E}[u(X_{T_k^n})|\mathscr{F}_n] \uparrow V_n.$$

Then $$\mathbb{E}[V_n|\mathscr{F}_{n-1}] = \lim_k \uparrow \mathbb{E}[u(X_{T_k^n})|\mathscr{F}_{n-1}] \leq V_{n-1}.$$

It follows from the preceding argument applied to the supermartingale V_n that

$$Ru(X_n) \leq V_n \qquad \forall n \geq 0$$

and hence

$$Ru(X_0) \leq V_0 = \text{ess sup}_{\mathscr{T}_0} \mathbb{E}[u(X_T)|\mathscr{F}_0].$$

Taking the expectation $\mathbb{E}_x$ gives

$$\begin{aligned} Ru(x) &\leq \mathbb{E}_x \lim \uparrow \mathbb{E}[u(X_{T_k^0})|\mathscr{F}_0] = \lim \mathbb{E}_x[u(X_{T_k^0})] \\ &\leq \sup_{T \in \mathscr{T}} \mathbb{E}_x[u(X_T)]. \end{aligned}$$

So, finally, $Ru(x) = \sup_{T\in\mathscr{T}} \mathbb{E}_x[u(X_T)]$, which proves the proposition. □

C.4 Computation of Reduced Functions with Respect to Semi-groups

Let (P_t) be a Feller semi-group on $(E, \mathscr{E})$ and let $(U_\lambda)_{\lambda>0}$ be the associated resolvent family and $((X_t), (\mathscr{F}_t^0), \mathbb{P}_x)$ be the associated Markov process, as in Section A.1.

C.4.1 Notation

The cone of bounded supermedian functions is

$$C_1 = \{u \in b\mathscr{E}_+;\ \lambda U_\lambda u \leq u \quad \forall \lambda > 0\}.$$

The cone of bounded semi-group-supermedian functions is

$$C_2 = \{u \in b\mathscr{E}_+;\ P_t u \leq u \quad \forall t > 0\}.$$

The cone of bounded excessive functions is

$$C_3 = \{u \in C_2; \lim_{t\to 0} P_t u = u\}.$$

For $u \in b\mathscr{E}_+$, we define

$$R_i(u) = \inf\{v \in C_i\,;\ v \geq u\}, \qquad i = 1, 2, 3.$$

Let $u \in b\mathscr{E}_+$ and let $R^{(\lambda)}u$ be the réduite (Definition C.1.1) with respect to the kernel λU_λ.

Proposition C.4.2

$$\forall u \in b\mathscr{E}_+, \quad R_1(u) = \sup_\lambda R^{(\lambda)}u \in b\mathscr{E}_+$$

Proof Let $\Gamma_\lambda = \{v \in b\mathscr{E}_+ : \lambda U_\lambda v \leq v\}$. It follows from

$$\mu U_\mu - \lambda U_\lambda = (\mu - \lambda)U_\lambda (I - \mu U_\mu), \qquad \lambda, \mu > 0$$

that Γ_λ decreases as λ increases. Hence $R^{(\lambda)}u$ increases with λ. Now, $w = \lim_{\lambda\to\infty} R^{(\lambda)}u$ is measurable and belongs to $\Gamma_\mu\ \forall\mu$, so $w \in \cap_\lambda \Gamma_\lambda = C_1$. Finally, $f \in C_1$ and $f \geq u$ clearly implies $f \geq w$, and so we have $w = R_1(u)$. □

Lemma C.4.3 The cone of bounded excessive functions with respect to the resolvent, *that is*,

$$C_4 = \{u \in C_1; \lim_{\lambda\uparrow\infty} \lambda U_\lambda u = u\},$$

is equal to C_3.

Proof (a) If $u \in C_3$, then

$$\lambda U_\lambda u(x) = \int_0^\infty \lambda e^{-\lambda t} P_t u(x)\, dt.$$

Since $\lim_{t\to 0} P_t u(x) = u(x)$, the fact that the measure $\lambda e^{-\lambda t}\, dt$ converges narrowly to δ_0 as $\lambda \to \infty$ gives $\lim_{\lambda\to\infty} \lambda U_\lambda u(x) = u(x)$ and so $u \in C_4$.

(b) We refer to Meyer [(1)] for the converse. □

Proposition C.4.4 Let $u = v - w$, with $v, w \in C_3$. Then

$$R_1(u) = R_2(u) = R_3(u).$$

Proof The réduite property (Corollary C.2.2) applied to the sub-Markov kernel λU_λ gives the réduite property for supermedian functions:

$$(*) \qquad \forall f, g \in C_1 \qquad f - R_1(f - g) \in C_1.$$

Indeed, using the notation of the proof of Proposition C.4.2, we have

$$f - R^{(\lambda)}(f - g) \in \Gamma_\lambda.$$

Letting λ tend to infinity on the left hand side, and using $\lambda < \mu \Rightarrow \Gamma_\mu \subset \Gamma_\lambda$ and Proposition C.4.2, we get

$$f - R_1(f - g) \in \Gamma_\lambda.$$

But $C_1 = \bigcap_\lambda \Gamma_\lambda$, so (*) is proved.

Now put $v_1 = v - R_1(v - w)$ and $v_2 = R_1(v - w)$, so that we have $v_1 \in C_1, v_2 \in C_1$ and $v_1 + v_2 = v \in C_3$. Then

$$v_1 \geq \lim_{\lambda \to \infty} \lambda U_\lambda v_1$$

$$v_2 \geq \lim_{\lambda \to \infty} \lambda U_\lambda v_2$$

and

$$v_1 + v_2 = \lim_{\lambda \to \infty} \lambda U_\lambda (v_1 + v_2) \qquad \text{(by Lemma C.4.3).}$$

But, by Lemma C.4.3, this implies

$$v_1 = \lim_{\lambda \to \infty} \lambda U_\lambda v_1 \quad \text{and} \quad v_2 = \lim_{\lambda \to \infty} \lambda U_\lambda v_2 \quad \text{and so} \quad v_1, v_2 \in C_3.$$

The proposition follows, since $C_1 \supset C_2 \supset C_3$ implies $R_1(u) \leq R_2(u) \leq R_3(u)$. □

Let $t \in \mathbb{R}_+^*$. For $u \in b\mathscr{E}_+$, we write $R_{(t)}u$ for the réduite of u with respect to the kernel P_t.

Definition C.4.5 A function $u \in b\mathscr{E}$ is said to belong to the domain of the extended generator of the semi-group $(P_t)_{t\geq 0}$ if there exists a Borel function v such that

$$U_p|v| < \infty \quad \text{and} \quad u = U_p(pu - v), \qquad p > 0.$$

In this case we write $u \in \mathscr{D}_e A$, and the function v is denoted by Au.

C.4.6 Remarks

- If $u, v \in b\mathscr{E}$, and $u = U_p(pu - v)$ for some $p \in]0, \infty[$, then $u = U_p(pu - v)$ for every $p \in]0, \infty[$.
- The function v is unique up to changes on a set of potential zero ($A \in \mathscr{E}$ is of potential zero if $U_p 1_A = 0$, $p > 0$). It is given explicitly by

$$v = \lim_{n\to\infty} n(u - nU_n u)$$

or by

$$v = \lim_{t\downarrow 0} \frac{1}{t}(P_t u - u).$$

These limits exist outside a set of potential zero [Mokobodzki, 1970].

- It is easy to prove (exercise) that for $u, v \in b\mathscr{E}$ the property

$$u = U_p(pu - v)$$

(for some $p > 0$ or for every $p > 0$) is equivalent to the property:

$$u(X_t) - u(X_0) - \int_0^t v(X_s)\,ds \text{ is an } (\mathscr{F}_t^0)\text{-martingale}$$
$$\text{under } \mathbb{P}_x \text{ for every } x.$$

The following result and the remainder of this section are inspired by Lamberton and Pagès [1991] to which we refer for a deeper study.

Proposition C.4.7 Let $u \in \mathscr{D}_e A$, with Au bounded.

$$\forall t, t' > 0, \qquad \|R_{(t)}u - R_{(t')}u\|_\infty \leq (t \vee t')\|Au\|_\infty.$$

Proof As noted in the preceding remarks, the process $u(X_t) - u(X_0) - \int_0^t Au(X_s)\,ds$ is a martingale that is bounded on every finite interval of time. By Proposition C.3.2,

$$R_{(t)}u = \sup_{S \in \mathscr{T}_t} \mathbb{E}_\cdot[u(X_S)] \quad \text{and} \quad R_{(t')}u = \sup_{S' \in \mathscr{T}_{t'}} \mathbb{E}_\cdot[u(X_{S'})],$$

where $\mathscr{T}_t$ (resp., $\mathscr{T}_{t'}$) is the set of bounded stopping times with values in the set $t\mathbf{N}$ (resp., $t'\mathbf{N}$).

If $S' \in \mathscr{T}_{t'}$, let $S = \inf\{s \in t\mathbf{N} : s \geq S'\}$. Clearly $S \in \mathscr{T}_t$ and $0 \leq S - S' \leq t$.

We have

$$\mathbb{E}_\cdot[u(X_S)] = \mathbb{E}_\cdot[u(X_{S'})] + \mathbb{E}_\cdot\left[\int_{S'}^{S} Au(X_s)\,ds\right]$$

and so

$$\sup_{S \in \mathscr{T}_t} \mathbb{E}_\cdot[u(X_S)] \geq \sup_{S' \in \mathscr{T}_{t'}} \mathbb{E}_\cdot[u(X_{S'})] - t\|Au\|_\infty.$$

The result follows easily. □

Proposition C.4.8 Let $u \in \mathscr{D}_e A$ with Au bounded and $t \in \mathbb{R}_+$. Then

(i) $R_1(u) = R_2(u) = R_3(u) = \lim_{n \uparrow \infty} \uparrow R_{(t/2^n)}u$;

(ii) $R_3(u) - R_{(t)}u \leq t\|Au\|_\infty$.

Proof The sequence $R_{(t/2^n)}u$ is increasing with n. For $v \in C_2, v \geq u$, we have $v \geq R_{(t/2^n)}u$ and hence $R_2(u) \geq \lim_{n \uparrow \infty} R_{(t/2^n)}u$.

By Proposition C.4.7, $\forall s, t > 0$

$$\lim_{n \uparrow \infty} R_{(t/2^n)}u = \lim_{n \uparrow \infty} R_{(s/2^n)}u.$$

We denote this limit by w. We have $w \in b\mathscr{E}_+$ and $\forall t\ P_t w \leq w$, so $w \in C_2$, and it follows that $w = R_2 u$, which proves part (i).

Part (ii) follows directly from

$$|R_{(t)}u - R_{(t/2^n)}u| \leq t\|Au\|_\infty$$

by letting n tend to infinity. □

The preceding results enable us to compute $R_3(u)$ for u in the closure of $\{f \in \mathscr{D}_e A; Af \in b\mathscr{E}\}$ under the uniform norm (this space contains $\mathscr{C}_0(E)$ since the domain of the strong generator of (P_t) is dense in $\mathscr{C}_0(E)$ and is included in $\{f \in \mathscr{D}_e A; Af \in b\mathscr{E}\}$).

If $u \in \{f \in \mathscr{D}_e A; Af \in b\mathscr{E}\}$ and $v \in \{f \in \mathscr{D}_e A; Af \in b\mathscr{E}\}$ is such that $|v - u| \leq \epsilon$, then $|R_3 v - R_3 u| \leq \epsilon$ (since, as seen in Section A.2.5, the constant functions are excessive). By proposition C.4.8 we can compute $R_3 v$ by choosing t such that $\|R_3 v - R_{(t)}v\|_\infty \leq \epsilon$ (this supposes we know a bound for $\|Av\|_\infty$). Finally, we can compute $R_{(t)}v$ by the réduite algorithm (cf. Proposition C.1.2 and Corollary C.1.3).

C.4.9 Application: The Réduite at Finite Horizon

Let $F: E \times \mathbb{R}_+ \to \mathbb{R}$ be $\mathscr{E} \otimes \mathscr{B}(\mathbb{R}_+)$-measurable and bounded, and put

$$Q_s F(x, t) = P_s F(x, s + t).$$

This defines a Feller semi-group on $E \times \mathbb{R}_+$.

We define

$$\tilde{C}_2 = \{F \geq 0 \text{ measurable bounded: } Q_s F \leq F \quad \forall s > 0\}$$
$$\tilde{C}_3 = \{F \in \tilde{C}_2 : \lim_{s \to 0} Q_s F = F\}$$

and let $\tilde{R}_2$ and $\tilde{R}_3$ be the réduite operators with respect to these cones.

Definition C.4.10 Let $u \in b\mathscr{E}$ and $T \in \mathbb{R}_+^*$. We put

$$R^T u = \tilde{R}_3(u 1_{[0,T]})(., 0).$$

Proposition C.4.11 Let $u \in \mathscr{D}_e A$ with $Au \in b\mathscr{E}$, and define $K_t u = u \vee P_t u$. Then,

$$R^T u = \lim_{n \uparrow \infty} \uparrow (K_{T/2^n})^{2^n} u$$

and

$$R^T u - (K_{T/2^n})^{2^n} u \leq \frac{T}{2^n} \|Au\|_\infty.$$

Proof Set $F(x,t) = u(x)1_{[0,T]}(t)$.

(a) Let us first remark that if $\tilde{R}_{(t)}G$ denotes the réduite of the function G with respect to the kernel Q_t, then we have

$$(\tilde{R}_{(T/2^n)}F)(.,0) = (K_{T/2^n})^{2^n} u.$$

Indeed, for $t = T/2^n$, we have

$$Q_t F(x,s) = P_t u 1_{[0,T]}(s+t)$$

and so

$$F \vee Q_t F = K_t u 1_{[0,T-t]}(s) + u 1_{]T-t,t]}(s).$$

Let $\tilde{K}_t F$ be the left-hand side. It follows by induction that

$$(\tilde{K}_t)^{2^n} F = (K_t)^{2^n} u 1_{\{0\}}(s) + (K_t)^{2^n-1} u 1_{]0,(T/2^n)]}(s) + \cdots$$

and this function is Q_t-excessive, which proves the assertion.

(b) Note that if (Y_t) is the process associated with the semi-group (Q_t), it is possible in proposition C.3.2 to restrict the sup to stopping times $\leq T$:

$$\tilde{R}_{(t)}F = \sup_{S \in \mathcal{F}_t, S \leq T} \mathbb{E}_.[F(Y_S)]$$

since

$$F(Y_S) \leq F(Y_{S \wedge T}).$$

(c) It follows that if $u \in \mathscr{D}_e A$ with $Au \in b\mathscr{E}$, we can apply the argument of Proposition C.4.7 to get

$$\|\tilde{R}_{(t)}F - \tilde{R}_{(t')}F\|_\infty \leq (t \vee t')\|Au\|_\infty.$$

Then, by Proposition C.4.8,

$$\tilde{R}_2 F = \tilde{R}_3 F = \lim_{n \uparrow \infty} \uparrow \tilde{R}_{(T/2^n)} F.$$

The result follows by taking the value at $t = 0$, and using the definition of $R^T u$ and part (a). □

D THE CARRÉ DU CHAMP OPERATOR

> There can be no doubt that the fountainhead, the inspiration, of all literature is the physical and social universe we live in, and the same is true about mathematics. Applied mathematics is bound to change, in part because the problems change and in part because the tools for their solution change. As we learn more and more about the world, and learn how to control some of it, we need to ask new questions, and as pure mathematics grows, sloughs off the excess, and becomes both deeper and simpler thereby, it offers applied mathematics new techniques to use. [Paul R. Halmos, 1981]

The idea of a carré du champ operator is derived from that of a generator of a Markov process. The notion of generator can be defined in several ways [Bouleau (1),(4); Dellacherie and Meyer, 1987], and the same holds for the carré du champ operator.

The formal relationship is as follows: If A is the generator of a Markov process, the carré du champ operator Γ is defined by

$$\Gamma(f, g) = A(fg) - fAg - gAf.$$

The definition of Γ involves the domain of A and the question of whether or not this domain is an algebra.

We shall make use of the extended generator, which has already been introduced in Definition C.4.5. What was proved in the framework of Feller processes can be extended to right processes [Getoor, 1975; Dellacherie and Meyer, 1987]. For the origin of the idea of carré du champ operator see Hirsch [1976/77], for the idea of carré du champ operator associated with a symmetric semi-group and its connection with the idea presented here, we refer to Bouleau and Hirsch [1991, Chapter I, Section 4].

Throughout this section, we consider a Feller process $(X_t)_{t \geq 0}$ with transition semi-group $(P_t)_{t \geq 0}$ and resolvent family $(U_p)_{p \geq 0}$. The notation is the same as in Section A.

D.1 The Extended Generator and Its Domain

The idea was introduced in the following form by P. A. Meyer [1976] in his works on Littlewood–Paley inequalities. Recall Definition C.4.5.

Definition D.1.1 The function u is said to belong to the domain of the extended generator, $\mathscr{D}_e A$ with $Au = v$ if

(i) $u \in b\mathscr{E}$

(ii) v is Borel and $U_p|v|$ is bounded (for some $p > 0$ or for all $p > 0$)

(iii) $u = U_p(pu - v)$ (for some $p > 0$ or for all $p > 0$).

Under assumptions (i) and (ii), condition (iii) is equivalent to

(iii′) $\forall x \in E \quad C_t^u = u(X_t) - u(X_0) - \int_0^t v(X_s)\,ds$ is an $(\mathscr{F}_t^0, \mathbb{P}_x)$ martingale $\forall x \in E$.

It follows that the function v is uniquely defined up to a set of potential zero (a set $F \in \mathscr{E}$ such that $U_p 1_F = 0$ for some $p > 0$ or for all $p > 0$).

The main importance of the idea of extended generator comes from the characterization of the cases when $\mathscr{D}_e A$ is an algebra: we shall use the following proposition [Meyer, 1976; Mokobodzki, 1970; Feyel, 1979; Bouleau (1)].

Proposition D.1.2 The following properties are equivalent:

(i) $\mathscr{D}_e A$ is an algebra.

(ii) For every law μ, for every $(\mathscr{F}_t^0, \mathbb{P}_\mu)$ square integrable martingale M, the measure $d\langle M, M\rangle_t$ on $\mathbb{R}_+$ is absolutely continuous with respect to dt.

(iii) C^2 functions operate on $\mathscr{D}_e A$: $\forall d \geq 1$ if $F: \mathbb{R}^d \to \mathbb{R}$ is C^2, and $u_1, \ldots, u_d \in \mathscr{D}_e A$ then $F(u_1, \ldots, u_d) \in \mathscr{D}_e A$.

(iv) There exists a non-affine function φ that is the difference of convex functions and operates on $\mathscr{D}_e A$.

We assume from now on that these conditions are satisfied. The Feller process is then said to be of *Lebesgue type* (because of property (ii)). In this case, $\Gamma(f, g)$ is defined for $f, g \in \mathscr{D}_e A$ by

$$\Gamma(f, g) = A(fg) - fAg - gAf. \tag{D.1.3}$$

The carré du champ operator enables us to calculate the bracket of some

martingales. We refer to Chapter 1 and to Dellacherie and Meyer [1980] for the stochastic calculus used below.

Proposition D.1.4 Let $f \in \mathscr{D}_e A$ and $g = Af$. The martingale

$$C_t^f = f(X_t) - f(X_0) - \int_0^t g(X_s)\, ds$$

has the bracket

$$\langle C^f, C^f \rangle_t = \int_0^t \Gamma(f,f)(X_s)\, ds.$$

Proof It is easy to see that C^f is locally square integrable: $\mathbb{E}_x[(C_t^f)^2] < +\infty \forall t \, \forall x \in E$. If we apply Ito's formula to the semi-martingale $Y_t = f(X_t)$, we get

$$f^2(X_t) = f^2(X_0) + \int_0^t 2f(X_s)g(X_s)\, ds + [C^f, C^f]_k$$

$$+ 2\int_{]0,t]} Y_{s-}\, dC_s^f. \tag{D.1.5}$$

by Proposition D.1.2. By the fact (Motoo's theorem) that an absolutely continuous additive functional is of the form $\int_0^t h(X_s)\, ds$ for some universally measurable function h, we have

$$\langle C^f, C^f \rangle_t = \int_0^t h(X_s)\, ds.$$

Then, by formula (D.1.5), the process

$$f^2(X_t) - f^2(X_0) - 2\int_0^t f(X_s)g(X_s)\, ds - \int_0^t h(X_s)\, ds$$

is a local martingale under $\mathbb{P}_x \forall x$. Hence, by property (iii′) of Definition D.1.1, we get $Af^2 = h + 2fg$ and thus $h = Af^2 - 2fAf$. The result is proved. □

D.2 Application to Hedging Strategies in Incomplete Markovian Markets

The idea of representing the evolution of stock values by stochastic processes has taken on new dimensions with the works of Black and Scholes, and Merton, in which the principle of pricing by simulation is introduced. (The word simulation here is taken in a specific sense.)

This is expressed mathematically by representing the contingent claim as a stochastic integral with respect to the underlying stock price. This is the object of the complete market assumption of Harrison and Pliska [1981]. Nevertheless, stochastic calculus enables us to study models in which the claim cannot be completely simulated.

Following Bouleau and Lamberton [1989], we take a Markovian framework and derive explicit formulas. We assume that the stock prices are functions of a Markov process. The carré du champ operator allows us to give an easily applicable form to the results.

The state of the market is represented by a Feller process (X_t) with filtration $(\mathscr{F}_t^0)$. We assume that the discounted underlying stock price S_t is an $(\mathscr{F}_t^0)$-martingale, which we write as

$$S_t = G(t, X_t),$$

where G satisfies suitable hypotheses [Bouleau and Lamberton, 1989], and that the contingent claim is European with discounted value of the form $H(S_T)$, where T is the exercise time.

A self-financing hedging strategy for the contingent claim creates a portfolio whose discounted value at time t is

$$V_t = V_0 + \int_{[0,t]} J_s \, dS_s,$$

where V_0 is equal to the initial value of the hedging portfolio and (J_t) is an $(\mathscr{F}_t^0)$-predictable process. The residue at time T is the lack of hedging

$$R_T = H(S_T) - V_T.$$

We have the following estimate of the variance of R_T:

$$\begin{aligned} \mathbb{E}(R_T^2) \geq\ & \mathbb{E}(V_0 - \mathbb{E}(H(S_T)|\mathscr{F}_0))^2 \\ & + \mathbb{E}\left[\int_0^T \left[\Gamma(F,F) - \frac{\Gamma(F,G)^2}{\Gamma(G,G)}\right](x, X_s)\, ds\right], \end{aligned} \tag{D.2.1}$$

where Γ is the carré du champ operator. Γ can normally be computed by the formula

$$\Gamma(F, G) = A(FG) - FAG - GAF,$$

where A is the generator of (X_t) operating on the x variable, and $F(t,x) = [P_{T-t}H(G(T, .))](x)$, (P_t) is the transition semi-group of (X_t).

In the estimate (D.2.1), equality is obtained for the unique hedging strategy that minimizes the quadratic risk $\mathbb{E}[\mathbb{R}_T^2]$, and is obtained by letting

$$V_0 = \mathbb{E}(H(S_T)|\mathscr{F}_0^0),$$

$$J_t^{(\text{optimal})} = (\Gamma(F, G)/\Gamma(G, G))(t, X_{t-}). \tag{D.2.2}$$

It is now natural to define the value of the claim at time 0 by $\mathbb{E}(H(S_T)|\mathscr{F}_0^0) = F(0, X_0)$ and at time t by $\mathbb{E}(H(S_T)|\mathscr{F}_t^0) = F(t, X_t)$.

The residue at time t can then be defined by

$$R_t = F(t, X_t) - V_t.$$

For the optimal strategy, the martingale (R_t) satisfies

$$\langle R, R\rangle_t = \int_0^t \left(\Gamma(F, F) - \frac{\Gamma(F, G)^2}{\Gamma(G, G)} \right)(s, X_s)\, ds,$$

which, by Doob's inequality, gives an estimate for the maximal residue during the optimal management of the portfolio:

$$\mathbb{E} \sup_{0 \le t \le T} R_t^2 \le 4 \int_0^t P_s \left(\Gamma(F, F) - \frac{\Gamma(F, G)^2}{\Gamma(G, G)} \right)(s, x)\, ds, \tag{D.2.3}$$

where $x = X_0$ is the starting point of the process (X_t).

These formulas extend to the case where there are several stocks (S_t with values in $\mathbb{R}^d$); see Bouleau and Lamberton [1989].

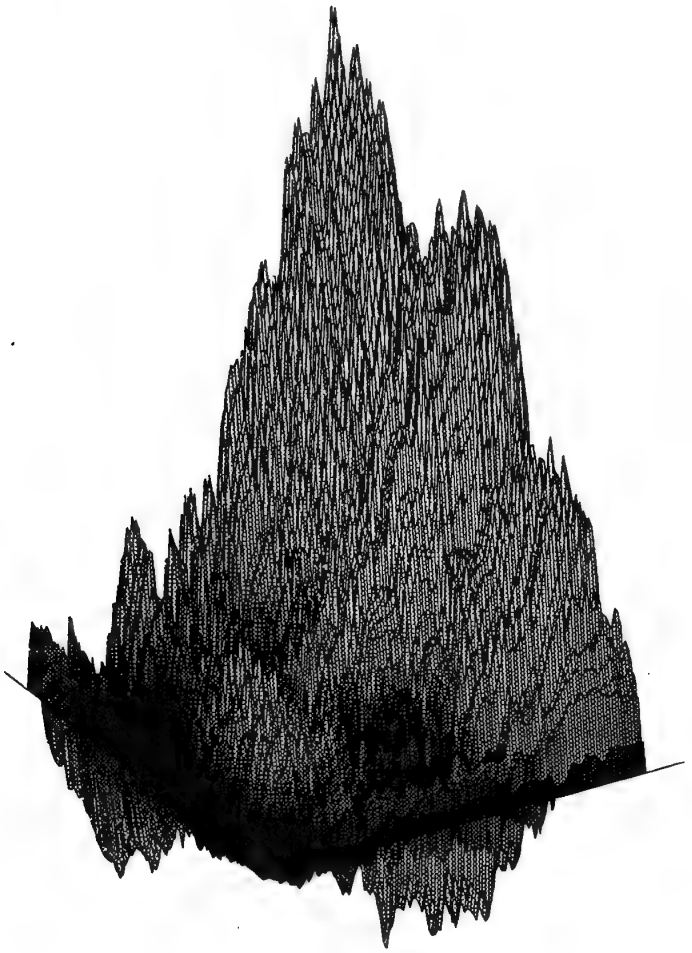


Figure 4.1 Recursively simulated Brownian sheet.

Appendix for Chapter 4

The Brownian sheet is a Gaussian process $X = (X_{s,t})$ which can be defined as follows.

Let (χ_n) be an C.O.N.S. of $L^2([0,1]^2)$, and let (g_n) be an i.i.d. sequence of standard Gaussian random variables. For $A \in \mathscr{B}([0,1]^2)$ the formula

$$Q(A) = \sum_{n=0}^{\infty} g_n \langle 1_A, \chi_n \rangle$$

defines a COSM (cf IIB.1) (Centred Orthogonal Stochastic Measure). Then $X_{s,t} = Q([0,s] \times [0,t])$. Putting $\psi_n(s,t) = \int_0^s \int_0^t \chi_n(\alpha,\beta)\, d\alpha\, d\beta, X$ is also defined by the series

$$X = \sum_{n=0}^{\infty} g_n \psi_n$$

which can be shown to converge in $\mathscr{C}([0,1]^2)$ a.s. Such a process can be recursively simulated by an extension of the method exposed in II.C.3. The idea is to use the conditional law of $X_{s+h,t+h}$ given

$$X_{s,t}, X_{s+h,t}, X_{s+2h,t}, X_{s,t+h}, X_{s+2h,t+h}, X_{s,t+2h}, X_{s+h,t+2h}, X_{s+2h,t+2h}$$

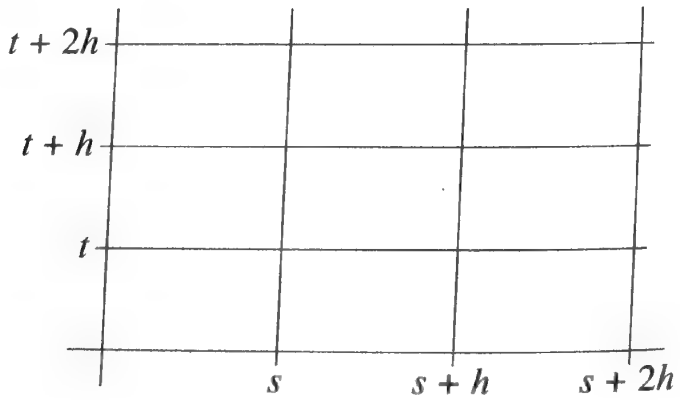
which can be obtained by a symbolic computer system using the fact that the vector

$$X_{s,t}, X_{s+h,t}, X_{s+2h,t}, X_{s,t+h}, X_{s+h,t+h}, X_{s+2h,t+h}, X_{s,t+2h}, X_{s+h,t+2h}, X_{s+2h,t+2h}$$

has the same law as the vector

$$\sqrt{st}N_1,$$
$$\sqrt{st}N_1 + \sqrt{th}N_2,$$
$$\sqrt{st}N_1 + \sqrt{th}N_2 + \sqrt{th}N_3,$$
$$\sqrt{st}N_1 + \sqrt{sh}N_4,$$
$$\sqrt{st}N_1 + \sqrt{th}N_2 + \sqrt{sh}N_4 + hN_5,$$
$$\sqrt{st}N_1 + \sqrt{th}N_2 + \sqrt{th}N_3 + \sqrt{sh}N_4 + hN_5 + hN_6,$$
$$\sqrt{st}N_1 + \sqrt{sh}N_4 + \sqrt{sh}N_7,$$
$$\sqrt{st}N_1 + \sqrt{th}N_2 + \sqrt{sh}N_4 + hN_5 + \sqrt{sh}N_7 + hN_8,$$
$$\sqrt{st}N_1 + \sqrt{th}N_2 + \sqrt{th}N_3 + \sqrt{sh}N_4 + hN_5 + hN_6 + \sqrt{sh}N_7 + hN_8 + hN_9$$

where $(N_1, N_2, N_3, N_4, N_5, N_6, N_7, N_8, N_9)$ are independent standard Gaussian random variables.



CHAPTER 5

Stochastic Differential Equations and Brownian Functionals

In this chapter we investigate the numerical resolution of stochastic differential equations and the computation of expectations of random variables defined on Wiener space.

A LIPSCHITZIAN STOCHASTIC DIFFERENTIAL EQUATIONS: ITO'S THEOREM

A.1 Notation and Lemmas

Let $(\Omega, \mathcal{F}_t, \mathbb{P})$ be a probability space with filtration $(\mathcal{F}_t)$. We will assume that martingales are continuous on the right with left limits. Let M be an $(\mathcal{F}_t)$-martingale.

A.1.1 Doob Inequality

Let p, q be such that $1 < p < +\infty$, and $1/p + 1/q = 1$. For all $t \in \mathbb{R}_+$,

$$\| \sup_{s \le t} |M_s| \|_p \le q \|M_t\|_p.$$

A.1.2 Burkholder Inequality

If $1 \le p < +\infty$, then $\forall t \in \mathbb{R}_+$

$$\| \sup_{s \le t} |M_s| \|_p \le 4p \| \langle M, M \rangle_t^{1/2} \|_p.$$

A.1.3 Gronwall Lemma

Let $u, v : \mathbb{R}_+ \to \mathbb{R}_+$ be continuous functions and let $c \ge 0$. If

$$v(t) \le c + \int_0^t v(s)u(s)\,ds \qquad \forall t \le T$$

then

$$v(t) \le c \exp \int_0^t u(s)\,ds \qquad \forall t \le T.$$

A.2 Lipschitz Hypotheses

A function f from $\mathbb{R}_+ \times \mathbb{R}^d$ to $\mathbb{R}^m$ satisfies the Lipschitz hypotheses if it is Borel and

$$\forall T \ge 0 \qquad \exists C_T \ge 0 \qquad \forall x, y \in \mathbb{R}^d \qquad \forall t \in [0,T]$$

$$|f(t,x)| \le C_T(1 + |x|)$$

and

$$|f(t,x) - f(t,y)| \le C_T|x - y|.$$

A.3 Ito's Theorem

Let $\sigma : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^{n\times d}$ and $b : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$ satisfy the Lipschitz hypotheses. We consider the stochastic differential equation

$$\text{(A.3.1)} \qquad \begin{cases} dX_t = \sigma(t, X_t)\,dB_t + b(t, X_t)\,dt \\ X_0 = x, \end{cases}$$

where (B_t) is a standard $(\mathscr{F}_t) - \mathbb{P}$-Brownian motion.

Let (U_t) and (V_t) be progressively measurable processes, and put (when this makes sense)

$$(\mathscr{J}U)_t = x + \int_0^t \sigma(s, U_s)\,dB_s + \int_0^t b(s, U_s)\,ds$$

and $\Delta_t(U,V) = \mathbb{E}[\sup_{s\le t} |U_s - V_s|^p]$ for $p \ge 2$.

We begin by looking at hypotheses under which $\mathscr{J}U$ is defined. Using $|a+b+c|^p \le 3^{p-1}(|a|^p + |b|^p + |c|^p)$, we have

$$\mathbb{E}|\sup_{s\le t}(\mathscr{J}U)_s|^p \le 3^{p-1}\left(|x|^p + \mathbb{E}\left[\sup_{s\le t}\left|\int_0^s \sigma(r,U_r)\,dB_r\right|^p + \sup_{s\le t}\left|\int_0^s b(r,U_r)\,dr\right|^p\right]\right)$$

and, by the Burkholder and Cauchy–Schwarz inequalities,

$$\le 3^{p-1}\left(|x|^p + (4p)^p\mathbb{E}\left(\int_0^t |\sigma(r,U_r)|^2\,dr\right)^{p/2} + \mathbb{E}\left(\int_o^t |b(r,U_r)|\,dr\right)^p\right)$$

$$\le 3^{p-1}\left(|x|^p + (4p)^p t^{(p/2)-1}\int_0^t \mathbb{E}|\sigma(r,U_r)|^p\,dr + t^{p-1}\int_0^t \mathbb{E}|b(r,U_r)|^p\,dr\right)$$

$$\le 3^{p-1}\left(|x|^p + ((4p)^p t^{(p/2)-1} + t^{p-1})C_T^p\int_0^t \mathbb{E}(1+\sup_{r\le s}|U_r|)^p\,ds\right).$$

Therefore, if U is an adapted continuous process such that $\forall t \ge 0, \mathbb{E}[\sup_{s\le t}|U_s|^p] < \infty$, we deduce that $\mathscr{J}U$ is well defined and, by Gronwall's lemma,

$$\mathbb{E}[\sup_{s\le T}|(\mathscr{J}U)_s|^p] \le C(p,T)$$

for a finite constant $C(p,T)$.

We now consider two continuous adapted processes U and V such that

$$\forall t \in \mathbb{R}_+, \mathbb{E}[\sup_{s\le t}|U_s|^p] < \infty \qquad \text{and} \qquad \mathbb{E}[\sup_{s\le t}|V_s|^p] < \infty.$$

The same chain of inequalities gives

$$\Delta_t(\mathscr{J}U, \mathscr{J}V) \le 2^{p-1}\mathbb{E}\left[\sup_{s\le t}\left|\int_0^s (\sigma(r,U_r)-\sigma(r,V_r))\,dB_r\right|^p + \sup_{s\le t}\left|\int_0^s (b(r,U_r)-b(r,V_r))\,dr\right|^p\right]$$

and, by the Burkholder and Cauchy–Schwarz inequalities,

$$\leq 2^{p-1}4^p p^p \mathbb{E}\left\{\left[\int_0^t |\sigma(r,U_r)-\sigma(r,V_r)|^2\,dr\right]^{p/2}\right\}$$
$$+ 2^{p-1}\mathbb{E}\left\{t^{p-1}\int_0^t |b(r,U_r)-b(r,V_r)|^p\,dr\right\}$$
$$\leq 2^{p-1}4^p p^p \mathbb{E}\left\{t^{(p/2)-1}\int_0^t |\sigma(r,U_r-\sigma(r,V_r)|^p\,dr\right\}$$
$$2^{p-1}\mathbb{E}\left\{t^{p-1}\int_0^t |b(r,U_r)-b(r,V_r)|^p\,dr\right\}.$$

Hence we have

(A.3.2) $$\Delta_t(\mathscr{F}U,\mathscr{F}V) \leq K(p,T)\int_0^t \Delta_s(U,V)ds.$$

Thus, putting $X^{(0)} = x$, $X^{(n+1)} = \mathscr{F}X^{(n)}$, we obtain by induction, for $t \leq T$,

$$\Delta_t(X^{(n)},X^{(n+1)}) \leq \Delta_T(X^{(0)},X^{(1)})\,\frac{K(p,T)^n t^n}{n!}.$$

Therefore

$$\sum_{n=1}^{\infty}\left\|\sup_{s\leq t}|X_s^{(n)}-X_s^{(n-1)}|\right\|_p < +\infty$$

and $\sum_{n=1}^{\infty}\sup_{s\leq t}|X_s^{(n)}-X_s^{(n-1)}| < +\infty$ a.s.

It follows that $X^{(n)}$ converges uniformly on compact sets a.s. and in L^p to a continuous process X. This implies that $\mathscr{F}X^{(n)}$ converges in the same sense to $\mathscr{F}X$, and so

$$X = \mathscr{F}X.$$

The uniqueness of the solution comes from Gronwall's lemma, since if U and V are solutions that are progressively measurable and

$$\forall t \geq 0,\ \mathbb{E}[\sup_{s\leq t}|U_s|^p] < \infty \qquad \text{and} \qquad \mathbb{E}[\sup_{s\leq t}|V_s|^p] < \infty,$$

we have, by formula (A.3.2), that

$$\Delta_t(U,V) \le C(p,T)\int_0^t \Delta_r(U,V)\,dr.$$

Therefore $\Delta_T(U,V) = 0$ and U and V are indistinguishable.

A more general uniqueness result can be proved using a stopping argument; see Revuz and Yor [1991] and Protter [1990].

Theorem A.3.3 Under the Lipschitz hypotheses on σ and b, equations (A.3.1) possess a unique solution $X = (X_t)$. The Picard iteration $X^{(n)}$ converges a.s. and in L^p:

$$\sup_{s\le t} |X_s^{(n)} - X_s| \to 0 \quad \text{a.s.} \qquad \text{and} \qquad \text{in} \quad L^p \quad p \in \quad [1,\infty[.$$

A.4 Two Properties of the Solution

Proposition A.4.1 For $1 \le p < +\infty$, and $t \le T$,

$$\|X_t\|_p \le C(p,T)(1 + |x|).$$

Proof We may assume that $p = 2q$, $q \ge 1$. For simplicity of the notation we suppose that the dimension is one. Applying Ito's formula to $|X_t|^{2q}$ gives

$$|X|_t^{2q} = x^{2q} + \int_0^t 2q|X|_s^{2q-1}(\sigma(s,X_s)dB_s + b(s,X_s)\,ds) + \frac{1}{2}\int_0^t 2q(2q-1)|X|_s^{2q-2}\sigma(s,X_s)^2\,ds.$$

We know that

$$\mathbb{E}\int_0^t |X|_s^{4q}\,ds \le T(\mathbb{E}\sup_{s\le t}|X_s|^{4q}) < +\infty.$$

Hence taking the expectation erases the stochastic integral and gives

$$\mathbb{E}|X|_t^{2q} \le x^{2q} + C(q,T)\int_0^t [(1 + \mathbb{E}|X|_s)^{2q}]\,ds$$
$$\le (1 + |x|)^{2q}C_1(q,T) + C_2(q,T)\int_0^t \mathbb{E}(|X_s|^{2q})\,ds,$$

and the result follows from Gronwall's lemma. □

Proposition A.4.2 For $1 \le p < +\infty$, $0 \le s, t \le T$

$$\|X_t - X_s\|_p \le c(p,T)(1+|x|)(t-s)^{1/2}.$$

Proof

$$\begin{aligned}\mathbb{E}|X_t - X_s|^p &\le 2^{p-1}\left[\mathbb{E}\left|\int_s^t \sigma(r,X_r)\,dB_r\right|^p + \mathbb{E}\left|\int_s^t b(r,X_r)\,dr\right|^p\right] \\ &\le 2^{p-1}4^p p^p (t-s)^{(p/2)-1}\int_s^t \mathbb{E}|\sigma(r,X_r)|^p\,dr \\ &\quad + 2^{p-1}(t-s)^{(p-1)}\int_s^t \mathbb{E}|b(r,X_r)|^p\,dr \\ &\le C(p,T)|t-s|^{p/2}(1+|x|)^p.\end{aligned}$$

□

B DISCRETIZATION OF SDEs

B.1 Euler Scheme: Convergence and Rate Estimates

Our aim is to solve (A.3.1) by a more effective method than Picard's iteration.

For $l \in \mathbf{N}^*$, let $t_k = k(T/l)$, $0 \le k \le l$ be a set of points partitioning $[0,T]$. We define $X_t^{(l)}$ on $]t_k, t_{k+1}]$ by

$$X_t^{(l)} = X_{t_k}^{(l)} + \int_{t_k}^t \sigma(t_k, X_{t_k}^{(l)})\,dB_s + \int_{t_k}^t b(t_k, X_{t_k}^{(l)})\,ds \tag{B.1.1}$$

with $X_0^{(l)} = x$.

It is easy to compute explicit expressions for the values of $X_t^{(l)}$ at the points of the partition. By induction on k we have $X_{t_k}^{(l)} \in L^p$, $1 \le p < +\infty$, and

$$\begin{aligned}X_t^{(l)} = x &+ \int_0^t \sum_{j=0}^{l-1} \sigma(t_i, X_{t_j}^{(l)}) 1_{[t_j, t_{j+1}[}(s)\,dB_s \\ &+ \int_0^t \sum_{j=0}^{l-1} b(t_i, X_{t_j}^{(l)}) 1_{[t_j, t_{j+1}[}(s)\,ds.\end{aligned}$$

Lemma B.1.2 For $1 \le p < +\infty$,

$$\sup_{t \le T} |X_t^{(l)}| \in L^p.$$

Indeed, if $Y_t = \sup_{s \le t} |X_s^{(l)}|$, then by the Burkholder and Cauchy–Schwarz inequalities and the Lipschitz hypotheses,

$$\mathbb{E} Y_t^p \le C \left[|x|^p + (t^{p-1} \vee t^{(p/2)-1}) \int_0^t \mathbb{E}(Y_s^p)\, ds \right].$$

The result follows by Gronwall's lemma.

We compute the difference $X_t^{(l)} - X_t$:

$$X_t^{(l)} - X_t = \int_0^t \sum_{j=0}^{l-1} (\sigma(t_j, X_{t_j}^{(l)}) - \sigma(x, X_s)) 1_{[t_j, t_{j+1}[}(s)\, dB_s$$
$$+ \int_0^t \sum_{j=0}^{l-1} (b(t_j, X_{t_j}^{(l)}) - b(x, X_s)) 1_{[t_j, t_{j+1}[}(s)\, ds.$$

We now assume the Hölder property:

(B.1.3) $\exists \alpha > 0 |\sigma(t,x) - \sigma(s,x)| + |b(t,x) - b(s,x)| \le C_T(1 + |x|)|t - s|^\alpha$.

Setting $Z_t = \sup_{s \le t} |X_s^{(l)} - X_s|$, we can write

$$\mathbb{E} Z_t^p \le C(p,T) \left[\mathbb{E} t^{(p/2)-1} \int_0^t \left| \sum_{j=0}^{l-1} (\sigma(t_j, X_{t_j}^{(l)}) - \sigma(s, X_s)) 1_{[t_j, t_{j+1}[}(s) \right|^p ds \right.$$
$$\left. + \mathbb{E} t^{p-1} \int_0^t \left| \sum_{j=0}^{l-1} b(t_j, X_{t_j}^{(l)} - b(s, X_s) 1_{[t_j, t_{j+1}[}(s) \right|^p ds \right].$$

For $s \in [t_j, t_{j+1}[$,

$$
\begin{aligned}
&|\sigma(t_j, X_{t_j}^{(l)}) - \sigma(s, X_s)| \\
&\quad \le |\sigma(t_j, X_{t_j}^{(l)}) - \sigma(t_j, X_{t_j})| + |\sigma(t_j, X_{t_j}) - \sigma(s, X_{t_j})| \\
&\qquad + |\sigma(s, X_{t_j} - \sigma(s, X_s)| \\
&\quad \le C[Z_s + (1 + |X_{t_j}|)(s - t_j)^{\alpha} + |X_s - X_{t_j}|]
\end{aligned}
$$

and so

$$
\mathbb{E}|\sigma(t_j, X_{t_j}) - \sigma(s, X_s)|^p \le C(p, T, x)\left[\mathbb{E}Z_s^p + \frac{1}{l^{\alpha p}} + \frac{1}{l^{p/2}}\right]
$$

(because $\mathbb{E}(1 + |X_j|^p) \le C(p, T, x)$ and $\mathbb{E}(|X_s - X_{t_j}|^p) \le C(s - t_j)^{p/2}$). Similarly, we have

$$
\mathbb{E}|b(t_j, X_{t_j}^{(l)}) - b(s, X_s)|^p \le C(p, T, x)\left[\mathbb{E}Z_s^p + \frac{1}{l^{\alpha p}} + \frac{1}{l^{p/2}}\right]
$$

and thus

$$
\mathbb{E}Z_t^p \le C(p, T, x)(t^{(p/2)-1})\left[\int_0^t \mathbb{E}Z_s^p \, ds + \frac{t}{l^{\alpha p}} + \frac{t}{l^{p/2}}\right].
$$

So by Gronwall's lemma,

$$
\mathbb{E}Z_t^p \le C(p, T)\frac{1}{l^{\beta p}} \qquad \text{with} \qquad \beta = \frac{1}{2} \wedge \alpha > 0,
$$

that is

$$
\| \sup_{t \le T} |X_t^{(l)} - X_t| \|_p \le \frac{C}{l^{\beta}}.
$$

Thus we obtain the following theorem:

Theorem B.1.4 Under the Lipschitz hypotheses and (B.1.3), the process $(X_t^{(l)})$ converges a.s. and in L^p uniformly on compact sets to the solution of the SDE.

It remains only to prove a.s. convergence. Choosing p such that $\beta p > 1$, we have

$$\mathbb{E}(\sup_{t\le T}|X_t^{(l)}-X_t|)^p \le \frac{C}{l^{\beta p}},$$

which shows that

$$\sum_l (\sup_{t\le T}|X_t^{(l)}-X_t|)^p$$

is integrable, and hence

$$\sup_{t\le T}|X_t^{(l)}-X_t| \to 0 \qquad \text{a.s.}$$

B.1.5 Remark

The approximating process $X^{(n)} = (X_t^{(n)})_{0\le t\le T}$ is easy to compute only at the points $t_k = kT/n$ of the partition.

If we ask for an approximation of the diffusion over the entire interval $[0,T]$, it is simplest to take a linear interpolation between the points $X_{t_k}^{(n)}$. This linearly interpolated scheme has almost the same rate of convergence as the preceding one:

As previously, we define for $t_k = kT/n, k = 0,\dots,n-1$,

$$\begin{cases} X_{t_{k+1}}^{(n)} = X_{t_k}^{(n)} + \sigma(t_k, X_{t_k}^{(n)})(B_{t_{k+1}} - B_{t_k}) + b(t_k, X_{t_k}^{(n)})(t_{k+1}-t_k) \\ \text{for } k = 0, \dots, n-1 \\ \text{and } X_0^{(n)} = x \end{cases}$$

and for $t \in [t_k, t_{k+1}[,\ k = 0,\dots,n-1$, put

$$\tilde{X}_t^{(n)} = X_{t_k}^{(n)} + \frac{t-t_k}{t_{k+1}-t_k}(X_{t_{k+1}}^{(n)} - X_{t_k}^{(n)}).$$

Then, under the Lipschitz hypotheses and assuming that (B.1.3) is satisfied with $\alpha \ge \frac{1}{2}$,

$$\| \sup_{t\in[0,T]} |X_t - \tilde{X}_t^{(n)}| \|_p \le C\sqrt{\frac{1+\log n}{n}} \qquad n \ge 1,$$

where C is a constant depending on p, T and the constants of the Lip-

schitz hypotheses and of (B.1.3). We refer to O. Faure [1992, Chapter IV] for a proof and extensions of the result.

B.2 Example: Wiener Iterated Integrals

Let $B_t = (B_t^1, \ldots, B_t^d)$ be a d-dimensional Brownian motion. Consider

$$I_N(t) = \int_{0<s_1<\cdots<s_N<t} dB_{s_1}^{i_1} \cdots dB_{s_N}^{i_N},$$

where the indices i_j are chosen, with eventual repetition, from $\{1, \ldots, d\}$. Let $0 < t_1 < \cdots < t_k < \cdots$ be a partition of $\mathbb{R}_+$ and let

$$I^1(t) = \int_0^t \sum_k 1_{[t_k, t_{k+1}[}(s)\, dB_s^{i_1}$$

$$I^2(t) = \int_0^t \sum_k I^1(t_k) 1_{[t_k, t_{k+1}[}(s)\, dB_s^{i_2}$$

$$\vdots$$

$$I^N(t) = \int_0^t \sum_k I^{N-1}(t_k) 1_{[t_k, t_{k+1}[}(s)\, dB_s^{i_N}.$$

By induction, we get

$$\text{(B.2.1)} \qquad I^N(t_{k+1}) = \sum_{j_1<j_2<\cdots<j_N<k} (B_{t_{j_1+1}}^{i_1} - B_{t_{j_1}}^{i_1})(B_{t_{j_2+1}}^{i_2} - B_{t_{j_2}}^{i_2}) \cdots (B_{t_{j_N+1}}^{i_N} - B_{t_{j_N}}^{i_N}).$$

Proposition B.2.2 $I^N(t)$ converges to $I_N(t)$ uniformly on compacts sets almost surely and in L^p as $\tau = \sup_k (t_{k+1} - t_k)$ tends to zero.

Proof If we set $X_t^1 = B_t^{i_1}$, $X_t^{k+1} = \int_0^t X_s^k dB_s^{i_k}$ the process $X_t = (X_t^1, \ldots, X_t^N)$ is a diffusion process solution of a Lipschitz SDE driven by the multidimensional Brownian motion B. The above discretization is that of Euler scheme. Putting $\tau = \sup_k (t_{k+1} - t_k)$, we get the estimate

$$\| \sup_{s \le T} |I_N(s) - I^N(s)| \|_p \le C(p, N, T)\sqrt{\tau}. \qquad \square$$

B.3 Example of a Multistep Scheme

The second-order Langevin equation

$$\ddot{X}(t) = -\gamma(t)\dot{X}(t) + b(t, X(t)) + \sigma(t, X(t))\dot{B}(t),$$

where $\dot{B}(t)$ is a white noise, describes the motion of a particle in a noise-perturbed force field with friction. It can be interpreted as the vector stochastic equation

$$(*) \qquad \begin{cases} dX_t = V_t\, dt \\ dV_t = -\gamma(t)V_t\, dt + b(t, X(t))\, dt + \sigma(t, X(t))\, dB_t \\ \qquad X_0 = x, \quad V_0 = v. \end{cases}$$

Let

$$\beta(t) = \exp \int_0^t \frac{\gamma(s)}{2}\, ds,$$

and choose a partition of $]0, T]$ into l intervals $]t_k, t_{k+1}]$. We assume that γ is differentiable and define

$$\begin{aligned} X_t^{(l)} = (\beta(t))^{-1} \Bigg[& x + \left(\frac{\gamma(0)}{2} x + v \right) t \\ & + \int_0^t \left(\int_0^s \sum_{j=0}^{l-1} \beta(t_j) \left[\frac{1}{4} (2\gamma'(t_j) + \gamma^2(t_j)) X_{t_j}^{(l)} \right. \right. \\ & \left. \left. + b(t_j, X_{t_j}^{(l)}) \right] 1_{[t_j, t_{j+1}[}(r) dr \right) ds \\ & + \int_0^t \left(\int_0^s \sum_{j=0}^{l-1} \beta(t_j) \sigma(t_j, X_{t_j}^{(l)}) 1_{[t_j, t_{j+1}[}(r)\, dB_r \right) ds \Bigg]. \end{aligned}$$

Theorem B.3.1 Suppose that σ and b satisfy the Lipschitz conditions and the Hölder condition (B.1.3). Suppose also that there exists $K > 0$ such that

$$|\gamma'(t) - \gamma'(s)| \le K|t-s|^{\alpha}.$$

Then for any $p \ge 2$,

$$\| \sup_{s \le T} |X_s^{(l)} - X_s| \|_p \le \frac{c}{l^{\alpha \wedge 1}}.$$

Proof (a) Suppose first that $\gamma = 0$. Then

$$X_t^{(l)} = x + vt + \int_0^t \left(\int_0^s \sum_{j=0}^{l-1} b(t_j, X_{t_j}^{(l)}) 1_{[t_j, t_{j+1}[}(r) dr \right) ds$$
$$+ \int_0^t \left(\int_0^s \sum_{j=0}^{l-1} \sigma(t_j, X_{t_j}^{(l)}) 1_{[t_j, t_{j+1}[}(r)\, dB_r \right) ds$$
$$X_t = x + vt + \int_0^t \left(\int_0^s b(r, X_r)\, dr + \int_0^s \sigma(r, X_r) dB_r \right) ds.$$

By the stochastic Fubini theorem, if $s \le t$,

$$X_t - X_s = v(t-s) + \int_0^t (t - s \vee r)[b(r, X_r)\, dr + \sigma(r, X_r)\, dB_r]$$

and as in Proposition A.4.2,

$$\mathbb{E}[|X_t - X_s|^p] \le 3^{p-1} \Bigg[|v|^p |t-s|^p$$
$$+ \mathbb{E}\left[\left| \int_0^t (t - s \vee r) \sigma(r, X_r)\, dB_r \right|^p \right]$$
$$+ \mathbb{E}\left[\left| \int_0^t (t - s \vee r) b(r, X_r)\, dr \right|^p \right] \Bigg]$$
$$\le C(p, T, x, v)|t-s|^p.$$

It also follows from the stochastic Fubini theorem that

$$X_t^{(l)} - X_t = \int_0^t (t-r)\left[\sum_{j=0}^{l-1}(b(t_j, X_{t_j}^{(l)}) - b(r, X_r))1_{[t_j,t_{j+1}[}(r)\,dr + \sum_{j=0}^{l-1}(\sigma(t_j, X_{t_j}^{(l)}) - \sigma(r, X_r))1_{[t_j,t_{j+1}[}(r)\,dB_r\right].$$

We put

$$Z_t = \sup_{s\le t} |X_s^{(l)} - X_s|.$$

Then

$$\mathbb{E}[Z_t^p] \le C(p,T,x,v)\left[\int_0^t \mathbb{E}(Z_s^p)\,ds + \frac{t}{l^{\alpha p}} + \frac{t}{l^p}\right];$$

by the Gronwall lemma,

$$\|Z_t\|_p \le \frac{C}{l^{\alpha\wedge 1}}.$$

(b) Suppose finally that γ satisfies the condition of the theorem, and let (X, V) be the solution of (*). We put

$$Y_t = \beta(t)X_t$$
$$W_t = \tfrac{1}{2}\gamma(t)\beta(t)X_t + \beta(t)V_t.$$

Then we have

$$dY_t = W_t dt$$
$$dW_t = \tfrac{1}{4}(2\gamma'(t) + \gamma^2(t))Y_t dt + \beta(t)[b(t,X_t)\,dt + \sigma(t,X_t)\,dB_t]$$

and so (Y_t, W_t) is the solution of a frictionless equation (*) with b and σ replaced by

$$\bar{b}(t,x) = \tfrac{1}{4}(2\gamma'(t) + \gamma^2(t))x + \beta(t)(b(t,(\beta(t))^{-1}x)$$
$$\bar{\sigma}(t,x) = \beta(t)\sigma(t,(\beta(t))^{-1}x).$$

It is easily seen that $\bar{b}$ and $\bar{\sigma}$ satisfy the Lipschitz conditions and the Hölder condition with $\alpha \wedge 1$, and we have the approximation result from (a):

$$\| \sup_{s \le T} |Y_s^{(l)} - Y_s| \|_p \le \frac{c}{l^{\alpha \wedge 1}}.$$

Since β is a positive continuous function and

$$Y_t^{(l)} = \beta(t) X_t^{(l)},$$

the proof of the theorem is complete. □

B.3.2 Comment

As is usual when dealing with second-order differential equations, our approximation scheme is multistep. In fact, with $h = T/l$, we have:

$$X_s^{(l)} = x$$

$$\beta(t_1) X_{t_1}^{(l)} - X_0^{(l)} = vh + \bar{b}_0 \frac{h^2}{2} + \bar{\sigma}_0 \int_0^{t_1} (t_1 - s)\, dB_s$$

$$\beta(t_{i+1}) X_{t_{i+1}}^{(l)} + \beta(t_{i-1}) X_{t_{i-1}}^{(l)} - 2\beta(t_i) X_{t_i}^{(l)} = (\bar{b}_{i-1} + \bar{b}_i) \frac{h^2}{2} + \bar{\sigma}_{i-1} \int_{t_{i-1}}^{t_i} (s - t_{i-1})\, dB_s + \bar{\sigma}_i \int_{t_i}^{t_{i+1}} (t_{i+1} - s)\, dB_s$$

where

$$\bar{b}_i = \frac{1}{4} (2\gamma'(t_i) + \gamma^2(t_i)) \beta(t_i) X_{t_i}^{(l)} + \beta(t_i) b(t_i, X_{t_i}^{(l)})$$

$$\bar{\sigma}_i = \beta(t_i) \sigma(t_i, X_{t_i}^{(l)}).$$

In the one-dimensional case, the random variables

$$\int_{t_{i-1}}^{t_i} (t_i - s)\, dB_s \qquad \text{and} \qquad \int_{t_{i-1}}^{t_i} (s - t_{i-1})\, dB_s$$

are Gaussian centered with variance $h^3/3$ and covariance $h^3/6$, so they can easily be simulated.

B.4 Do Not Use the Shift Method Along the Sample Paths

Suppose that, by a discretization method similar to those in this section, we were able to replace the problem of computing the expectation of a Brownian functional by that of computing the expectation of a functional of the form

$$F = f(B_{t_1}, B_{t_2}, \ldots, B_{t_k}, \ldots, B_T),$$

where $0 < t_1 < \ldots < t_{n-1} < T$ partitions $[0, T]$.

The functional F is finite-dimensional, and the random variables $B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_T - B_{t_{n-1}}$ are independent Gaussian with known variance. It is therefore easy to simulate F on $([0,1]^n, \mathscr{B}([0,1]^n), dx^n)$.

If n is large, as explained in Chapter 2, Section C.5, quasi-Monte Carlo methods are no longer applicable and the only available methods are the classical Monte Carlo and shift methods.

There is no difficulty with the Monte Carlo method but, as we have mentioned, we must draw all the variables $B_{t_i}, i = 1, \ldots, n$ independently at each step.

Concerning the shift method, we have seen in Chapter 3, Section A.7 that it is useful not only in infinite-dimensional but also in finite-dimensional cases. But its application can be carried out in several ways (cf. Chapter 3, Section A.7) yielding different rates of convergence.

One of these ways is particularly bad: we suppose for simplicity that $t_{i+1} - t_i = T/n \; \forall i$, and consider the approach that involves changing the i.i.d. sequence

$$B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_{n-1}} - B_{t_{n-2}}, B_T - B_{t_{n-1}}$$

into the sequence

$$B_{t_2} - B_{t_1}, B_{t_3} - B_{t_2} \ldots, B_T - B_{t_{n-1}}, \sqrt{\frac{T}{n}}\, g,$$

where g is a standard Gaussian variable independent of the preceding variables. In such a method, the new discretized path is very close to the preceding one at every step, and, as the increment $t_{i+1} - t_i = T/n$ tends to zero

the path remains just as close! Practically, it can be tested by experiment, that the rate worsens as n increases [Ben Alaya, 1992].

To overcome this difficulty, the successive points of Wiener space given by the approximate Brownian paths yielded by the shift must be far away from each other in the Wiener space. It is possible, for example, to simulate the Brownian path recursively (cf. Chapter 3, Section C.3) and to shift when the random function is called: each recursively simulated approximated Brownian path is obtained from a random function shifted one step.

It is also possible to expand the Brownian motion in series, either with Haar functions or with other bases of $L^2[0, T]$. Here again, there are several ways to implement this, depending on the chosen method of indexing.

B.4.1 Remark

(1) As we explained in Sections D.5 and F.4, the accuracy of the discretization we get from expanding the Brownian motion in a series may depend on the sample path (cf. also 0. Faure and J. Gaines [1992]). In this case the shift method uses an infinite-dimensional product space $[0, 1]^{\mathbb{N}}$ as explained in Chapter 3, Section A.5.

(2) What we have said above about Brownian motion can also be said about processes with independent increments and Markov processes: do not apply the shift method on an iterative simulation because the procedure worsens as the discretization becomes sharper.

C IRREGULARITY OF BROWNIAN FUNCTIONALS

In order to discuss the value of the preceding approximations, and especially of the almost sure approximation, we must first investigate the regularity of typical Brownian functionals.

Let B_t be a standard real Brownian motion defined on the Wiener space $\Omega = \mathscr{C}_0([0, 1], \mathbb{R})$ equipped with the uniform topology and the Wiener measure $\mathbb{P}$. (The subscript $_0$ in $\mathscr{C}_0([0, 1])$ means that elements of $\mathscr{C}_0([0, 1], \mathbb{R})$ vanish at zero.)

C.1 Lack of Continuity of Stochastic Integrals

Proposition C.1.1 Let $f \in L^2([0, 1])$. If the random variable $F(\omega) = \int_0^1 f(s)\, dB_s$ possesses a continuous version, then there exists a bounded measure μ such that $\mu(\{0\}) = 0$ and $f(s) = \mu(]s, 1])$ Lebesgue a.s.

For the proof we need the following lemma.

Lemma C.1.2 For any nonempty open subset G of $\mathscr{C}_0([0, 1], \mathbb{R})$, $P(G) > 0$.

Proof of the Lemma (a) Suppose first that B is a standard one-dimensional Brownian motion on $\mathbb{R}_+$ and for $a > 0$, $0 < \alpha < \pi/2a$, and $0 \le t < \infty$, put

$$T = \inf\{s \ge 0 : |B_s| \ge a\}$$

$$Z_t = \cos \alpha B_t \exp \frac{\alpha^2}{2}\, t.$$

As it is the real part of a complex continuous martingale, Z is a real continuous martingale. It follows from the optional sampling theorem that, for any $t \ge 0$,

$$\mathbb{E}[Z_{T\wedge t}] = \mathbb{E}[Z_0] = 1.$$

We remark that

$$\cos \alpha B_{T\wedge t} \ge \cos\, \alpha a > 0$$

Letting $t \to \infty$ we obtain

$$\mathbb{E}\left[\exp \frac{\alpha^2}{2}\, T\right] = (\cos\, \alpha a)^{-1}.$$

Letting $\alpha \to \pi/2a$ we observe that T cannot be bounded a.s., and hence for any $t \ge 0$, $\mathbb{P}(T > t) > 0$.

(b) We now return to the Wiener space $\Omega = \mathscr{C}_0([0,1], R)$. From (a) we deduce that for any open ball $B(0,a)$ in Ω we have $\mathbb{P}(B(0,a)) > 0$. Now let $h \in L^2([0,1])$ and put

$$g(t) = \int_0^t h(s)\, ds, \qquad 0 \le t \le 1.$$

It follows from the Cameron–Martin–Girsanov theorem that the continuous process $(B_t - g(t), 0 \le t \le 1)$ is a Brownian motion with respect to the new probability measure

$$\tilde{\mathbb{P}}(A) = \mathbb{E}\left[1_A \exp\left\{\int_0^1 h(s)\, dB_s - \tfrac{1}{2}\int_0^1 (h(s))^2\, ds\right\}\right] \qquad A \in \mathscr{E}.$$

As $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are equivalent, and

$$\tilde{\mathbb{P}}(B(g,a)) = \mathbb{P}(B(0,a)) > 0,$$

we deduce that $\mathbb{P}(B(g,a)) > 0$. Since the Cameron–Martin space is dense in Ω, the proof is complete. □

Proof of the Proposition: Let $\hat{F}$ be a continuous function on Ω such that

$$\hat{F} = \int_0^1 f(s)\,dB_s \quad \text{a.s.}$$

The function $\hat{F} : \mathscr{C}_0([0,1],\mathbb{R}) \to \mathbb{R}$ is such that if (B_t^1) and (B_t^2) are two independent Brownian motions and α belongs to $[-1,1]$, then

$$\hat{F}(\alpha B^1 + \sqrt{1-\alpha^2}B^2) = \alpha\hat{F}(B^1) + \sqrt{1-\alpha^2}\hat{F}(B^2) \quad \text{a.s.}$$

Both sides are continuous maps from $\mathscr{C}_0([0,1],\mathbb{R}) \times \mathscr{C}_0([0,1],\mathbb{R})$ to $\mathbb{R}$, and are, therefore, everywhere equal. Thus

$$\hat{F}(\alpha x + \sqrt{1-\alpha^2}y) = \alpha\hat{F}(x) + \sqrt{1-\alpha^2}\hat{F}(y)$$

for all $x, y \in \mathscr{C}_0([0,1],\mathbb{R})$ and $\alpha \in [-1,1]$. Taking $y = 0$ gives $\hat{F}(0) = 0$ and $\hat{F}(ax) = a\hat{F}(x)$ for $a \in [-1,1]$, and hence for all $a \in \mathbb{R}$. So

$$\hat{F}(ax + by) = a\hat{F}(x) + b\hat{F}(y) \quad \forall x, y \in \mathscr{C}_0([0,1],\mathbb{R}) \quad \forall a, b \in \mathbb{R}.$$

Thus $\hat{F}$ is linear and continuous and there exists a bounded measure μ such that $\mu(\{0\}) = 0$ and $\hat{F}(\omega) = \mu(\omega)\forall\omega \in \Omega$. Integrating by parts, the equality $\mu(\omega) = \int_0^1 \omega(s)\mu(ds)$ gives

$$\mu(\omega) = \int_0^1 \mu(]s,1])\,dB_s \quad \text{a.s.}$$

It follows that $f(s) = \mu(]s,1])$ for Lebesgue almost every s. □

Corollary C.1.3 Let B^1 and B^2 be independent Brownian motions, and let G be a Borel version of $\int_0^1 B_s^1 dB_s^2$. Then the set $\{\omega_1 : G(\omega_1,\cdot)$ is continuous$\}$ is negligible.

Proof This follows from the proposition since $\{\omega_1 : s \to \omega_1(s)$ has bounded variation$\}$ is negligible.

C.2 Non-Riemann-Integrable Stochastic Integrals

We will show the irregularity of the iterated integral

$$F(\omega) = \int_0^1 B_s^1 dB_s^2 \left(= \int_0^1 \omega_1(s)\, d\omega_2(s) \right)$$

defined on the two-dimensional Wiener space $(\Omega, \mathscr{F}, m) = (\Omega_1, \mathscr{F}_1, m_1) \times (\Omega_2, \mathscr{F}_2, m_2)$ with $\Omega_i = \mathscr{C}_0([0,1], \mathbb{R})\, i = 1, 2$.

This integral can be considered as a generic example of functionals of diffusion processes solutions of an Ito's type SDE. Indeed, on the one hand, as we shall see later (Section E.4), these diffusion processes can be written as the sum of a series of iterated integrals, and on the other hand the approximate continuity at smooth points of the Ito map, which defines a diffusion process (Stroock–Varadhan theorem, cf. Ikeda and Watanabe [1981, Theorem 8.2, p. 432]) can be proved from that of F alone.

Proposition C.2.1 Let Ω be endowed with the uniform topology. For every nonempty open set G and every $a \in \mathbb{R}$,

$$\mathbb{P}\{\omega \in G : F(\omega) > a\} > 0$$

and

$$\mathbb{P}\{\omega \in G : F(\omega) < a\} > 0.$$

Proof We introduce the process (Lévy's area)

$$S(t) = \tfrac{1}{2}\left[\int_0^t \omega_1(s)\, d\omega_2(s) - \int_0^t \omega_2(s)\, d\omega_1(s) \right]$$

and the radial process

$$r(t) = \sqrt{(\omega_1(s))^2 + (\omega_2(s))^2}.$$

It follows from Ito's formula that

$$\tfrac{1}{2}r(t)^2 = \int_0^t r(s)\,d\xi(s) + t,$$

where

$$\xi(t) = \int_0^t \frac{\omega_1(s)}{r(s)}\,d\omega_1(s) + \int_0^t \frac{\omega_2(s)}{r(s)}\,d\omega_2(s).$$

The bivariate martingale $(S(t), \xi(t))$ satisfies:

$$\langle S, S\rangle_t = \tfrac{1}{4}\int_0^t r(s)^2\,ds$$
$$\langle S, \xi\rangle_t = 0$$
$$\langle \xi, \xi\rangle_t = t.$$

Thus, by Knight's theorem on the representation of continuous martingales as time changed Brownian motion, if $K(t)$ is the inverse of the increasing process $t \to \frac{1}{4}\int_0^t r(s)^2\,ds$, then

$$(\eta_t, \xi_t) = (S(K(t)), \xi(t))$$

is a bivariate Brownian motion. In particular, the process $(S(K(t)))_{t\geq 0}$ is independent of the process $(\xi(t))_{t\geq 0}$.

Since r^2 is the solution to a stochastic differential equation, where σ is Hölder continuous with exponent $1/2$ and b is constant, we observe [Yamada and Watanabe, 1971] that r is adapted to the $\mathbb{P}$-completed filtration generated by the driving Brownian motion ξ. Therefore $(S(K(t)))_{t\geq 0}$ is also independent of r.

Lemma C.2.2 Let

$$h(t) = (h_1(t), h_2(t)) = \left(\int_0^t u_1(s)\,ds, \int_0^t u_2(s)\,ds\right)$$

with $u_1, u_2 \in L^2([0,1])$. Then

$$\forall a \in \mathbb{R}, \forall \delta > 0,\ \mathbb{P}\{F(\omega + h) > a, \|\omega\|_\infty < \delta\} > 0.$$

Proof of the Lemma We have

$$F(\omega+h)=\int_0^1 \omega_1(s)\,d\omega_2(s)+\int_0^1 h_1(s)\,d\omega_2(s)+\int_0^1 \omega_1(s)\,dh_2(s)+\int_0^1 h_1(s)\,dh_2(s).$$

We remark that

$$\int_0^1 \omega_1(s)\,d\omega_2(s)=\tfrac{1}{2}\omega_1(1)\omega_2(1)+S(1)$$

and

$$\tfrac{1}{2}\omega_1(1)\omega_2(1)\geq -\tfrac{1}{4}r^2(1).$$

Moreover,

$$\int_0^1 h_1(s)\,d\omega_2(s)=h_1(1)\omega_2(1)-\int_0^1 \omega_2(s)u_1(s)\,ds$$

$$\geq -|h_1(1)|(r^2(1)\vee 1)-\left(\int_0^1 r^2(s)\,ds\right)^{1/2}\cdot\left(\int_0^1 u_1^2(s)\,ds\right)^{1/2}$$

$$\int_0^1 \omega_1(s)\,dh_2(s)\geq -\left(\int_0^1 r^2(s)\,ds\right)^{1/2}\left(\int_0^1 u_2^2(s)\,ds\right)^{1/2}$$

so that

$$F(\omega+h)\geq S(1)+H(r),$$

where

$$H(r) = -\tfrac{1}{4}r^2(1) - |h_1(1)|(r^2(1) \vee 1) - \left(\int_0^1 r^2(s)\,ds\right)^{1/2} (\|u_1\|_{L^2} + \|u_2\|_{L^2}) + \int_0^1 h_1(s)u_2(s)\,ds.$$

Hence

$$\mathbb{P}[F(\omega + h) > a, \|\omega\|_\infty < \delta] \geq \mathbb{P}^{\mathbb{P}}\left[\eta\left(\tfrac{1}{4}\int_0^1 r^2(s)\,ds\right) + H(r) > a; \|\omega\|_\infty < \delta\right]$$

and this last probability is strictly positive, because η is independent of r and the function

$$f(\alpha, \beta) = \mathbb{P}\{\eta(\alpha) - \beta > a\}$$

is > 0 for $\alpha > 0$ and $\beta \in \mathbb{R}$. So, this probability can be written as

$$\mathbb{E}\left[1_{\{\sup_{0\leq s\leq 1} r(s) < \delta\}} f\left(\tfrac{1}{4}\int_0^1 r(s)^2\,ds, \tfrac{1}{4}r^2(1)\right)\right]$$

and the result follows from the fact that $\mathbb{P}\{\sup_{0\leq s\leq 1} r(s) < \delta\} > 0$. □

Proof of Proposition C.2.1 (end) Let G be an open set in Ω. The set of Cameron–Martin functions $H = \{h : h(t) = \int_0^t u(s)\,ds, u \in L^2([0,1], \mathbb{R}^2)\}$ is dense in Ω. Let $h \in H$, h in G, and let $\mathbb{P}_h$ be defined by

$$\mathbb{P}_h = \exp\left[\int_0^1 u_1(s)\,d\omega_1(s) + \int_0^1 u_2(s)\,d\omega_2(s) - \tfrac{1}{2}\int_0^1 |u(s)|^2\,ds\right].\mathbb{P}.$$

$(\omega(t) - h(t))_{t\geq 0}$ is a standard Brownian motion under $\mathbb{P}_h$. An application of Lemma C.2.2 gives the result. □

In particular, by the proposition, if φ is any nonconstant bounded continuous function from $\mathbb{R}$ to $\mathbb{R}$, the random variable $\varphi(\int_0^1 \omega_1(s)\,d\omega_2(s))$ is not Riemann-integrable.

C.2.3 Remark
Lévy's area $S(t) = \frac{1}{2}[\int_0^t \omega_1(s)\,d\omega_2(s) - \int_0^t \omega_2(s)\,d\omega_1(s)]$ can be simulated (in the sense of Chapter 3, Section A.2) by a continuous random variable on a different probability space: If $(B_t^{(1)}, B_t^{(2)}, B_t^{(3)})_{t\geq 0}$ is a standard Brownian motion with values in $\mathbb{R}^3$, then $S(t)$ has the same law as

$$B^{(1)}_{1/4\int_0^t [(B_s^{(2)})^2+(B_s^{(3)})^2]\,ds} \tag{C.2.4}$$

by the property quoted after Lemma C.2.2. Furthermore, we can see by computing the characteristic function that $S(t)$ has the same law as the random variable

$$B_1^{(1)}\sqrt{\frac{1}{4}\int_0^t (B_s^{(2)})^2 + (B_s^{(3)})^2\,ds}. \tag{C.2.5}$$

If (B_t) is a standard Brownian motion, the law of the random variable $\int_0^t B_s^2\,ds$ is known. This was historically one of the first applications of the formulas for absolutely continuous change of probability measures on Wiener space [Cameron and Martin, 1945].

We have

$$\mathbb{E}\left[\exp\left\{-\int_0^t \lambda \tan[\lambda(s-t)]B_s\,dB_s - \frac{1}{2}\int_0^t \lambda^2 \tan^2[\lambda(s-t)]B_s^2\,ds\right\}\right] = 1.$$

However, by Ito's formula

$$\begin{aligned}
&-\lambda\int_0^t \tan[\lambda(s-t)]B_s\,dB_s \\
&\quad = \frac{\lambda^2}{2}\int_0^t \frac{B_s^2}{\cos^2[\lambda(s-t)]}\,ds + \frac{\lambda}{2}\int_0^t \tan[\lambda(s-t)]\,ds \\
&\quad = \frac{\lambda^2}{2}\int_0^t \{1+\tan^2[\lambda(s-t)]\}B_s^2\,ds + \frac{1}{2}\log|\cos(\lambda t)|.
\end{aligned}$$

Thus, substituting, we get

$$\mathbb{E}\left[\exp\left(\frac{\lambda^2}{2}\int_0^t B_s^2\,ds\right)\right] = \frac{1}{\sqrt{|\cos\lambda t|}},$$

and so, by analytic continuation,

$$\mathbb{E}\left[\exp\left(-\frac{\lambda^2}{2}\int_0^t B_s^2\,ds\right)\right] = \frac{1}{\sqrt{\cosh \lambda t}}.$$

Returning to (C.2.4) and (C.2.5), we get

$$\mathbb{E}e^{iuS(t)} = \mathbb{E}\left[\exp\left(-\frac{u^2}{2}\,\frac{1}{4}\int_0^t [(B_s^{(1)})^2 + (B_s^{(2)})^2]\,ds\right)\right]$$
$$= \frac{1}{\cosh \dfrac{ut}{2}}$$

and so $S(t)$ has a density given by

$$\frac{1}{t\cosh \pi x/t}.$$

Un autre problème, dont j'ai commencé l'étude en 1939, et dont j'ai donné la solution complète dans ma communication au Second Berkeley Symposium on Probability and Statistics (1950), est relatif à l'aire $S(t)$ comprise entre l'arc $A(0)A(t)$ et sa corde, c'est à dire puisque $A(0)$ est en O, l'aire balayée par le vecteur OA quand le point A décrit cet arc.

Il est naturel d'inscrire dans l'arc $A(0)A(t)$ une ligne polygonale $A_0A_1 \ldots A_n$ à côtés tous très petits, et de remplacer l'arc par cette ligne pour avoir une approximation de $S(t)$. On introduit ainsi des sommes de termes de la forme $\frac{1}{2}R_\nu\xi_\nu\sqrt{dt_\nu}$, R_ν désignant la distance A_0A_ν et $\xi_\nu\sqrt{dt_\nu}$ désignant la hauteur du triangle $OA_{\nu-1}A_\nu$ issue du sommet $A_{\nu-1}$; ξ_ν est une variable gaussienne réduite, indépendante de R_ν. Une telle somme est alors de la forme $\xi\sqrt{\Sigma R_\nu^2}$, ξ étant toujours une variable gaussienne réduite. A la limite, on obtient pour $S(t)$ la forme

$$U\xi, \quad \text{où} \quad U^2 = I + J, I = \sum_0^t X^2(u)\,du, J = \sum_0^t Y^2(u)\,du,$$

I, J et U étant indépendants (donc aussi U et ξ).

Ce résultat est indiqué dans mon mémoire de 1939, où j'ai aussi formé une équation aux dérivées partielles vérifiée par la fonction de répartition conjointe de $R(t)$ et $S(t)$. Mais je ne l'avais pas intégrée, et ce n'est que plus tard que j'ai pu, et même de deux manières différentes, déterminer celle de $R(t)$.

La première utilise la f.c. $\varphi(z)$ de I, obtenue en 1944 par Cameron et Martin. Il est facile d'en déduire celle de U^2, qui est $\varphi^2(z)$, et celle de $S(t) = U\xi$, qui est $1/\cosh(tz/2)$. C'est une loi qui était déja connue. Elle est naturellement a.c. et sa densité de probabilité est $f(x) = 1/(t\cosh(\pi x/t))$.[1] [Paul Lévy, 1970]

C.3 Other Topologies on Wiener Space

We might guess that the discontinuity of the functional $F(\omega) = \int_0^1 \omega^1(s)\,d\omega^2(s)$ comes from the fact that uniform topology on Wiener space is too weak and has to be reinforced.

Nevertheless the following result can be shown [Sugita, 1989]. Let Ω and H be as in the proof of proposition C.2.1, and let B be a Banach space such that H is dense in B, $B \subset \Omega$, $\mathbb{P}(B) = 1$. Then there is no version $\tilde{F}$ of F (i.e., $\tilde{F} = F$ $\mathbb{P}$-a.e.) such that $\tilde{F}|_B$ is continuous.

We refer to Sugita's article for a characterization of Wiener multiple integrals G for which such a Banach space B exists on which G is continuous.

C.4 Importance of L^p Estimates

Let us consider a SDE of Ito's type, given by equation (A.3.1) with regular coefficients. It is known [McShane, 1974; Doss, 1977; Sussmann, 1978]

[1]Another problem, which I began working on in 1939, and gave the complete solution to in my communication to the Second Berkely Symposium on Probability and Statistics (1950), is concerned with the area $S(t)$ bounded by the arc $A(0)A(t)$ and its chord; in other words, since $A(0)$ is at O, the area swept out by the line OA as the point A describes this arc.

It is natural to inscribe in the arc $A(0)A(t)$ a polygonal line $A_0A_1\ldots A_n$, with very small sides, and to replace the arc with this line so as to get an approximation of $S(t)$. We also introduce sums of the form $\frac{1}{2}R_\nu S_\nu\sqrt{dt_\nu}$, where R_ν denotes the distance A_0A_ν from the corner $A_{\nu-1}$; ξ_ν is a reduced Gaussian variable, independent of R_ν. Such a sum is then of the form $\xi\sqrt{\Sigma R_\nu^2}$, where ξ is still a reduced Gaussian variable. In the limit, we obtain for $S(t)$ the form

$$U\xi, \quad \text{where} \quad U^2 = I + J, I = \sum_0^t X^2(u)\,du, J = \sum_0^t Y^2(u)\,du,$$

where I, J and U are independent (and so U and ξ are as well). This result is given in my 1939 treatise, where I also give an equation involving partial derivatives, proved using the joint distribution function of $R(t)$ and $S(t)$. But I did not integrate it, and was only later able, in two different ways, to determine that of $R(t)$. The first uses the cumulative function $\phi(z)$ of I, obtained in 1944 by Cameron and Martin. From this it is easy to deduce that of U^2, which is $\phi^2(z)$, and that of $S(t)$, which is $1/\cosh(tz/2)$. This was already known. It is naturally a.c. and its probability density is $f(x) = 1/(t\cosh(\pi x/t))$.

that if the Frobenius commutativity condition

$$\forall x \in \mathbb{R}^n \quad \sigma'(x)[\sigma(x)u]v = \sigma'(x)[\sigma(x)v]u \qquad u, v \in \mathbb{R}^d$$

is satisfied, then a version of the solution can be chosen such that the Ito mapping

$$(x, \omega) \to X^x(\omega) \in \mathscr{C}([0, 1], \mathbb{R}^n)$$

is locally Lipschitz. In particular, this occurs when $d = 1$. In this case, whenever φ is a bounded continuous function from $\mathscr{C}([0, 1], \mathbb{R}^n)$ to $\mathbb{R}$, the random variable $\varphi(X^x)$ is Riemann-integrable and its expectation can be computed by simulation with an almost-sure approximation scheme, like that of Theorem B.1.4. There is a large literature on this subject; see especially Mc Shane [1974], Clark [1980], Pardoux and Talay [1985], Milstein [1974], Sussman [1978] and O. Faure [1990].

Now, as we have seen for a particular case in Sections C.1, C.2, and C.3, if the commutativity condition fails, then functionals like $\varphi(X^x)$ are no longer Riemann-integrable and their expectations cannot be computed by direct simulation.

The general method is therefore to use L^p-approximation results to find a continuous random variable F_ϵ such that

$$|\mathbb{E}F_\epsilon - \mathbb{E}\varphi(X^x)| \leq \frac{\epsilon}{2},$$

and F_ϵ can generally be defined on a finite-dimensional space. We can then compute $\mathbb{E}F_\epsilon$ by simulation with an accuracy of $\epsilon/2$, which gives the desired result up to ϵ. The computation of $\mathbb{E}F_\epsilon$ can be performed by classical Monte Carlo, by quasi-Monte Carlo, or by an ergodic transform as is used in the shift method.

C.5 Sequential Integration by Picking Smooth Functions in the Wiener Space

Under the expression "sequential integration" we group together all the numerical methods of computation of the integral of a function f as the limit of average of its values at a sequence of points (ξ_n):

$$\int f \, d\mu = \lim_{N \uparrow \infty} \frac{1}{N} \sum_{n=1}^{N} f(\xi_n).$$

In the finite-dimensional case ($f : [0,1]^s \to \mathbb{R}, s \in \mathbf{N}^*$) these methods are well-tried, for the class of functions f to which they apply, as well as for their rate of convergence. They can be classified in three main families: Monte Carlo, quasi-Monte Carlo (low-discrepancy sequences), and shift [Bouleau, 1993].

As already discussed, in the infinite-dimensional case, and in particular on the Wiener space $(W, \mathscr{W}, \mathbb{P}) = (\mathscr{C}_0([0,1], \mathbb{R}^r), \mathscr{B}(\mathscr{C}_0([0,1], \mathbb{R}^r), \mathbb{P})$ which will be here our main framework, the extension of these methods immediately meets some difficulties:

(a) A lot of interesting random variables for modelization (stochastic integrals, solutions of SDE's, etc.) are defined up to almost-sure equality. So, if (ξ_n) is a sequence of points in the Wiener space, the pointwise value $f(\xi_n)$ is not defined for such f's.

(b) These random variables are very irregular and their Borel versions are not $\mathbb{P}$-Riemann integrable. For example, for the bidimensional Brownian motion $\omega = (\omega^1, \omega^2)$ the random variable $f(\omega) = \int_0^1 \omega_s^1 \, d\omega_s^2$ is such that for every nonempty open set G and for every $a \in \mathbb{R}$

$$\mathbb{P}[G \cap \{f > a\}] > 0 \qquad \text{and} \qquad \mathbb{P}[G \cap \{f < a\}] > 0,$$

and for every Borel version $\tilde{F}$ of F, the set of continuity points of $\tilde{F}$ is empty.

(c) Even when f is continuous, since a typical Brownian path is essentially a limit concept, the value $f(\omega)$ is computable only when f is simple (i.e., simulatable in the sense of Section D below).

Fortunately, the problem is not completely hopeless because of the approximate continuity property discovered by Stroock and Varadhan [1972]; see also Hargé [1993]:

Let X_t be the solution of the SDE in Stratonovitch sense

$$\begin{cases} dX_t^i = \displaystyle\sum_{k=1}^{r} \sigma_k^i(X_t) \circ dB_t^k + b^i(X_t)\, dt \\ X_0 = x, \end{cases}$$

where the coefficients σ_k^i, b^i are $\mathscr{C}^3$, bounded with bounded derivatives.

$$\begin{cases} dX_t^i = \displaystyle\sum_{k=1}^{r} \sigma_k^i(X_t) \circ dB_t^k + b^i(X_t)\, dt \\ X_0 = x \end{cases}$$

Let also $Y_t = Y_t(\varphi)$ the solution of the ordinary differential equation

$$\begin{cases} dY_t^i = \sum_{k=1}^{r} \sigma_k^i(Y_t)\dot{\varphi}(t)\,dt + b^i(Y_t)\,dt \\ Y_0 = x \end{cases}$$

for $\varphi \in \mathscr{C}^\infty([0,1], \mathbb{R}^r)$, $\varphi(0) = 0$.

Then

$$\mathbb{P}\left[\sup_{t\in[0,1]} |X_t - Y_t(\varphi)| > \epsilon \,\middle|\, \sup_{t\in[0,1]} |\omega(t) - \varphi(t)| < \delta\right] \to 0$$

when $\delta \to 0$.

By this result, it is possible to hope that for some sequence of smooth functions (φ_n) "dense" enough in $(W, \mathscr{W}, \mathbb{P})$,

$$\mathbb{E}X_1 = \lim_{N\uparrow\infty} \frac{1}{N} \sum_{n=1}^{N} Y_1(\varphi_n).$$

See in this direction Bouleau [(12)].

D SIMULATABLE FUNCTIONALS

We consider first the problem of computing the expectation of, say, a stopping time $T = \inf\{n \geq 0 : X_n \in A\}$ of a Markov chain (X_n). If T is integrable, then we can apply the general method above, and we can in general define a bounded stopping time T_ϵ such that $|\mathbb{E}T_\epsilon - \mathbb{E}T| \leq \epsilon/2$, and then compute $\mathbb{E}T_\epsilon$ by simulating the chain. But, as we have seen, since T is a finite stopping time, such an approximation is not necessary and it becomes tractable to compute $\mathbb{E}T$ directly by averaging the values of T on independent samples. In this section, we study in more detail the random variables which can be pointwise simulated directly. We begin with discrete probability spaces.

D.1 Simulatable Functionals on $\{0,1\}^N$

Let $(x_n)_{n\in N}$ be the coordinate mappings of $\{0,1\}^N$ onto each factor. As usual, we define the σ-fields $\mathscr{A}_n = \sigma(x_m, m \leq n)$ and the product probability $\mathbb{P} = (\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1)^{\otimes N}$ on $\mathscr{A}_\infty$.

Definition D.1.1 A random variable F from $\{0,1\}^N$ to $\mathbb{R}$ is said to be *simple* if there exists a $\mathbb{P}$-almost surely finite $(\mathscr{A}_n)$-stopping times T such that F is $\mathscr{A}_T$-measurable.

In fact, we must often assume that F is bounded (see Chapter 3, Section A.3). Since $F = \sum_{n=0}^{\infty} F \cdot 1_{T=n}$, the law of F is discrete:

$$\mu = \sum_{k=1}^{\infty} p_k \delta_{a_k}.$$

It can be shown that an $(\mathscr{A}_n)$-stopping time S and an $\mathscr{A}_S$-measurable random variable G with law μ can be constructed such that $\mathbb{E}S$ is minimal. For $x \in [0,1]$, let $\{x\}$ be the fractional part of x, and put $\nu(x) = \sum_{n=0}^{\infty} \{2^n x\}/2^n$. Then the minimal value of $\mathbb{E}S$ (finite or infinite) is [Knuth and Yao, 1976]

$$\sum_{k=1}^{\infty} \nu(p_k).$$

In general, however, the laws of the simple random variables to be simulated are not known. This is precisely why the simulation is useful.

D.2 Simulatable Functionals on $[0,1]^N$

We now adopt the hypothesis that it is possible to pick from [0, 1] a sequence of points independently and according to Lebesgue measure. We use this for reasoning about the simulation algorithms, as does, for example, L. Devroye [1986].

Let $(U_n)_{n\geq 0}$ be the coordinate mappings of $[0,1]^N$, let $\mathscr{F}_n = \sigma(U_m, m \leq n)$, and let $\mathbb{P} = dx^{\otimes N}$. We will consider random variables F defined on $[0,1]^N$ which are $\mathscr{F}_T$-measurable for some $(\mathscr{F}_n)$-stopping time T that is $\mathbb{P}$-almost surely finite. But which regularity must we demand on F to get an interesting definition?

We must, at least, suppose that F is $\mathbb{P}$-Riemann integrable (on $[0,1]^N$ equipped with the product topology) so that, outside a negligible set, the value of F at a point x can be approximated by an approximate knowledge of x. However, it is not difficult to define quite explicitly a Borel mapping χ from [0, 1] into $[0,1]^N$ such that χ is continuous outside a negligible set and that the image of the measure dx is $dx^{\otimes N}$: Let

$$x = \sum_{k=0}^{\infty} \frac{x_k}{2^{k+1}} \qquad x_k \in \{0,1\}$$

be the canonical dyadic expansion of $x \in [0,1]$, and define

$$y = \chi(x) \quad \text{to be} \quad y = (y_n)_{n\geq 0} \quad \text{with} \quad y_n = \sum_{k=0}^{\infty} \frac{x_{k+(n+k)(n+k+1)/2}}{2^{k+1}}.$$

Thus if we suppose only that F is $\mathbb{P}$-Riemann integrable, it is possible to return to $[0,1]$ by considering $F \circ \chi$. But, since the digits of y_n depend on too faraway digits of x, this is in practice useless for the simulation of functionals of random processes.

So we must ask more regularity on F than just Riemann-integrability. It seems reasonable to take in consideration the following definition:

Definition D.2.1 A random variable F from $[0,1]^N$ to $\mathbb{R}^d$ is said to be *simple* if there is an $(\mathscr{F}_n)$-stopping time T, $dx^{\otimes N}$-almost surely finite such that the sets $\{T = n\}$ have negligible boundaries, F is $\mathscr{F}_T$-measurable and, for every n, F is continuous in the interior of $\{T = n\}$.

As before, we must often suppose that F is also bounded. It follows from the definition that F is continuous on a set of measure 1 and is therefore Riemann-integrable if it is bounded.

D.2.2 Example

Let X be a random variable with values in $\mathbb{R}^d$ and continuous (or Riemann integrable) density f on $\mathbb{R}^d$. If X is simulated by the rejection method with a random variable Y with continuous density g such that $f \leq kg$, and Y is simulated by $Y = \varphi(x), x \in [0,1]$, with φ continuous, then we can write

$$T = \inf\{2n+1 : kU_{2n+1}\, g \circ \varphi(U_{2n}) \leq f \circ \varphi(U_{2n})\}$$

and

$$X = \varphi(U_{T-1}).$$

T is a stopping time with geometric law, the sets $\{T = n\}$ have negligible boundaries, and X is simple in the sense of Definition D.2.1

More generally,

Definition D.2.3 Given a probability measure μ on $(\mathbb{R}^N, \mathscr{B}(\mathbb{R}^N))$ and a Polish space $(E, \mathscr{E})$, we say that a measurable mapping G from

$(\mathbb{R}^N, \mathscr{B}(\mathbb{R}^N))$ into $(E, \mathscr{E})$ is *simple* if there is a stopping time S for the filtration $(\mathscr{G}_n)$ of the coordinates (X_n) of $\mathbb{R}^N$, μ-almost surely finite, such that the boundaries of the sets $\{S = n\}$ are negligible and G is $\mathscr{G}_n$-measurable and continuous on the interior of $\{S = n\}$ for every n.

D.3 Regularity Properties of Simulatable Functionals

If the stopping time in the definition is not just almost surely finite but is everywhere finite, the laws of simple random variables depend continuously on the perturbations of the random number generator and depend analytically if the stopping time possesses an exponential moment. We take first the case of $\{0,1\}^N$ and suppose that the generator yields independent bits with law $p\delta_1 + (1-p)\delta_0$ for $p \in [0,1]$:

Proposition D.3.1

(i) Let F be simple on $\{0,1\}^N$ with respect to a finite stopping time T. Then the distribution function $H_p(t) = \mathbb{P}_p\{F \le t\}$ under probability $\mathbb{P}_p = (p\delta_1 + (1-p)\delta_0)^{\otimes N}$ is continuous in p, for $p \in [0,1]$.

(ii) If, in addition, there is an $\alpha > 0$ such that

$$\mathbb{E}_{1/2}[(1+2\alpha)^T] < +\infty$$

then $H_p(t)$ is analytical in p in an open set of the complex plane containing $]\frac{1}{2} - \alpha, \frac{1}{2} + \alpha[$.

Proof (i) With the above notation, we have

$$H_p(t) = \sum_{n=0}^{\infty} \mathbb{P}_p[\{F \le t\} \cap \{T = n\}].$$

$\mathbb{P}_p[\{F \le t\} \cap \{T = n\}]$ is a polynomial (of degree $\le n+1$) that is positive on $[0,1]$. Therefore $p \to H_p(t)$ is l.s.c. and, by the same argument, so is $p \to 1 - H_p(t) = \mathbb{P}_p[F > t]$, and hence $p \to H_p(t)$ is continuous.

(ii) Let $M_n = \prod_{k=0}^{n}[1 + (2p-1)(2x_k - 1)]$ be the martingale, which is the density of $\mathbb{P}_p$ with respect to $\mathbb{P}_{1/2}$ on $\mathscr{A}_n$.

We have

$$H_p(t) = \sum_{n=0}^{\infty} \mathbb{E}_{1/2}[M_n 1_{\{F \le t\}} 1_{\{T=n\}}].$$

Putting $q = p - \frac{1}{2}$, the absolute value of the expectation

$$\mathbb{E}_{1/2}[M_n 1_{\{F\le t\}} 1_{\{T=n\}}] = \mathbb{E}_{1/2}\left[\prod_{k=0}^{n}(1+2q(2x_k-1))1_{\{F\le t\}}1_{\{T=n\}}\right]$$

is bounded by $(1+2|q|)^{n+1}\mathbb{P}_{1/2}[\{T=n\}]$.

$H_p(t)$ is therefore the sum of a series of holomorphic functions on $\{z : |z-\frac{1}{2}| < \alpha\}$ uniformly convergent in every compact set of this domain and the result follows. □

These properties extend to $\mathcal{F}_S$-measurable functionals and, in particular, to simple functionals on $[0,1]^N$: let h be a bounded Borel function on $[0,1]$ such that $\int_0^1 h(x)\,dx = 0$ and put $C = \|h\|_\infty$. Suppose that our generator yields independent points distributed according to the law $(1+\lambda h(x))\cdot dx$ for $\lambda \in [-1/C, 1/C]$. Then we have the following proposition:

Proposition D.3.2

(i) Let F be an $\mathcal{F}_S$-measurable function from $[0,1]^N$ to $\mathbb{R}$, where S is a $(\mathcal{F}_n)$-stopping time which is almost surely finite for the probability measures $\mathbb{P}_\lambda = [(1+\lambda\cdot h(x))\,dx]^{\otimes N}$ for $\lambda \in [-a,a]$ $(a \le 1/C)$. Then the distribution function $H_\lambda(t) = \mathbb{P}_\lambda\{F \le t\}$ is continuous in λ on $[-a,a]$.

(ii) Suppose moreover $\mathbb{E}_0[(1+\alpha C)^S] < +\infty$ for some $\alpha > 0$. Then $H_\lambda(t)$ is analytical in $\{\lambda : |\lambda| < \alpha\}$.

The proof is similar to the previous one.

D.4 Simulatable Functionals on a General Probability Space

In this section we propose a general definition of simulatable functionals.

> La théorie, énumération jamais complète d'un ensemble de règles qui s'amorcent l'une l'autre, n'est donc pas le produit de la froide raison, mais de l'instinct qui travaille silencieusement, allumant au foyer intérieur un feu mystérieux dont la raison ne fait que recueillir les cendres brûlantes.[2] [André Lhote, 1919]

[2]The theory, which is the never complete enumeration of a set of rules which initiate each other, is not the product of cool reasoning. It results from the quietly working instinct, that lights the inner hearth a mysterious fire, the burning ashes of which are only gathered by reason.

Consider a Polish space W equipped with its Borelian σ-field $\mathcal{W}$ and a probability measure m. For example, $(W, \mathcal{W}, m)$ could be the Wiener space or the canonical space of some stochastic process.

As the preceding discussion shows, the topology on the space W is in general too strong to be used in simulation. Indeed, to be used effectively, the topology must be such that a fundamental system of neighborhoods of a point is composed of sets described by a finite number of rational numbers, that is, by an integer. On the Wiener space, for instance, we could take simple convergence on rational numbers (cf. Section D.5 below).

We define a presentation of $(W, \mathcal{W}, m)$ as a copy of its measurable structure equipped with a weaker topology:

Definition D.4.1 Let $(W, \mathcal{W}, m)$ be a Polish space. A presentation of $(W, \mathcal{W}, m)$ is a mapping φ from W into $\mathbb{R}^N$ such that

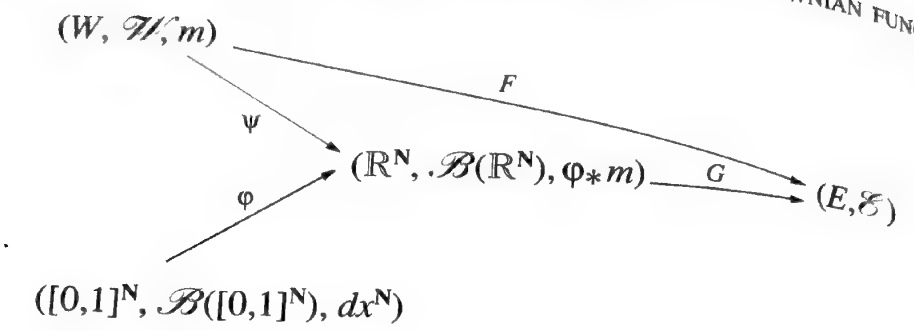
- φ is an isomorphism of measurable spaces between $(W, \mathcal{W}, m)$ and $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N), \mu)$ [i.e., $\mu = \varphi_* m$ and $\exists N, m(N) = 0$, $\exists N^1, \mu(N^1) = 0$) such that φ is one to one and bimeasurable from $W \backslash N$ onto $\mathbb{R}^N \backslash N^1$].
- φ is continuous.

Next we must simulate progressively the coordinate process $X = (X_n)_{n \geq 0}$ of $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N), \mu)$ so that if we are given a simple functional G on $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N), \mu)$ we can simulate G too.

Definition D.4.2 Let $X = (X_n)_{n \geq 0}$ be the coordinates of $\mathbb{R}^N$. We say that $\psi : [0,1]^N \to \mathbb{R}^N$ is a progressive simulation of $X = (X_n)_{n \geq 0}$ under μ if

- ψ is measurable from $([0,1]^N, \mathcal{B}([0,1])^N)$ to $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$ and $\psi_*(dx^{\otimes N}) = \mu$.
- There exists an increasing sequence of $(\mathcal{F}_n)$-stopping $T_k, k \geq 0$, $dx^{\otimes N}$-almost surely finite such that $\forall k \geq 0$ the map $(X_0, \ldots, X_k) \circ \psi$ from $[0,1]^N$ into $\mathbb{R}^{k+1}$ is $\mathcal{F}_{T_k}$-measurable.

Finally, we say that a functional F defined on $(W, \mathcal{W}, m)$ with values in the Polish space $(E, \mathcal{E})$ is an s-functional if we can find a presentation φ of $(W, \mathcal{W}, m)$, a simple functional G from $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N), \varphi_* \mu)$ into $(E, \mathcal{E})$ and a progressive simulation ψ such that $F = G \circ \varphi$.



The preceding definitions will become clearer by their application in the context of Wiener space.

D.5 The Case of Wiener Space

Let $(W, \mathscr{W}, m)$ be the space of standard Brownian motion vanishing at 0, where $W = \mathscr{C}_0([0,1], \mathbb{R}^d)$, $\mathscr{W}$ is the Borelian σ-field on W equipped with the uniform convergence, and m is the Wiener measure.

We look at several presentations of the Wiener space.

D.5.1 Dyadic Presentation

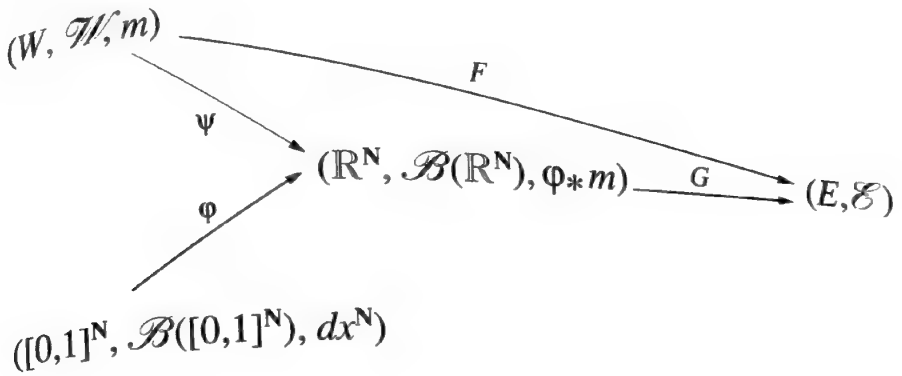
As before, let X_n be the coordinate mappings of $\mathbb{R}^N$, and put

$$X_0 \circ \varphi(w) = w(1)$$
$$X_{2^n+k} \circ \varphi(w) = w\left(\frac{2k+1}{2^{n+1}}\right) \qquad k = 0, 1, \ldots, 2^n - 1; \quad n \geq 0.$$

This defines a continuous mapping φ from W to $\mathbb{R}^N$ (equipped with the product topology). The coordinates (X_n) constitute a Gaussian process under the probability measure $\varphi_* m$ and the law of the vector $(X_0, \ldots, X_{2^n-1})$ is a permutation of the law of $(w(1/2^n), \ldots, w(k/2^n), \ldots, w(1))$. φ is clearly an isomorphism of measurable spaces.

It is easy here to give a progressive simulation ψ of the process $X = (X_n)$ on the space $([0,1]^N, \mathscr{B}[0,1]^N, dx^N)$. This consists of simulating $X_{2^n}, X_{2^n+1}, \ldots, X_{2^{n+1}-1}$ given $X_0, \ldots, X_{2^n-1}$, which can be done by Paul Lévy's celebrated recursive definition of Brownian motion of (cf. Chapter 3, Section C).

Thus we have again the diagram



We shall now see that there are simple functionals G, which define interesting s-functionals $F = G \circ \varphi$.

D.5.2 Example

For $\epsilon > 0$, we define

$$T_\epsilon = \inf \left\{ 2^n : \sup_{1 \le k \le 2^n} \left| w\left(\frac{k}{2^n}\right) - w\left(\frac{k-1}{2^n}\right) \right| \le \epsilon \right\}$$

and denote by $L_{2^n}(w, t)$ the piecewise linear dyadic approximation of order n of $w \in W$. So, $w \to L_{2^n}(w, t)$ is a mapping with values in $\mathscr{C}_0([0, 1], \mathbb{R}^d)$ equipped with its Borel sets. This is a Polish space.

Let us put $F_\epsilon = L_{T_\epsilon}$.

Then F_ϵ is an s-functional: it factorizes as $F_\epsilon = G_\epsilon \circ \varphi$, and we have the following lemma:

Lemma D.5.3 G_ϵ is a simple functional (Definition D.2.3).

Proof It is easily seen that we must prove the properties of Definition D.2.3 for F_ϵ and T_ϵ when W is equipped with the topology of simple convergence on dyadic points and with the σ-fields $\mathscr{H}_{2^n} = \sigma(w(k/2^n), k = 1, \ldots, 2^n)$.

(a) Clearly T_ϵ is an $(\mathscr{H}_{2^n})$-stopping time m-almost surely finite, and $\mathscr{F}_\epsilon$ is $\mathscr{H}_{T_\epsilon}$-measurable.

(b) Let us put $\Delta_{2^n}(W) = \sup_{1 \le k \le 2^n} |\omega(k/2^n) - \omega((k-1)/2^n)|$.

We remark that the set

$$A_\epsilon = \{ w : \Delta_{T_\epsilon}(w) < \epsilon \}$$

is open for the topology of simple convergence on dyadic points. Indeed, if $w_0 \in A_\epsilon$ and if the $T_\epsilon(w_0)$ points

$$\left(w\left(\frac{k}{T_\epsilon(w_0)} \right), \qquad k = 1, \ldots, T_\epsilon(w_0) \right)$$

are sufficiently close to the respective points

$$\left(w_0\left(\frac{k}{T_\epsilon(w_0)} \right), \qquad k = 1, \ldots, T_\epsilon(w_0) \right),$$

then $T_\epsilon(w) = T_\epsilon(w_0)$ and hence $L_{T_\epsilon}(w)$ is closed to $L_{T_\epsilon}(w_0)$. It follows that F_ϵ is continuous on the interior of $\{T_\epsilon = 2^n\}$ and it can be proved that the boundary of $\{T_\epsilon = 2^n\}$ is negligible. □

In the same way the functionals

$$F_\epsilon^1 = L_{T_\epsilon^1} \quad \text{where} \quad T_\epsilon^1 = \inf\left\{ 2^n : \sum_k \left| w\left(\frac{k}{2^n} \right) - w\left(\frac{k-1}{2^n} \right) \right| \geq \frac{1}{\epsilon} \right\}$$

$$F_\epsilon^2 = L_{T_\epsilon^2} \quad \text{where} \quad T_\epsilon^2 = \inf\left\{ 2^n : \sum_k \left| w\left(\frac{k}{2^n} \right) - w\left(\frac{k-1}{2^n} \right) \right|^2 \in [1-\epsilon, 1+\epsilon] \right\}$$

are s-functionals.

D.5.4 Example

We take again the example of Section C:

$$F(w) = \int_0^1 w_s^1 dw_s^2.$$

Notice the following simply proved facts.

(i) $\mathbb{E}[\int_s^t w_u^1 dw_u^2 | \, w_s = (a_1, b_1), w_t = (a_2, b_2)] = ((a_1 + a_2)/2)(b_2 - b_1)$.

(ii) $\mathbb{E}[(\int_s^t w_u^1 dw_u^2 - ((a_1 + a_2)/2)(b_2 - b_1))^2 | w_s = (a_1, b_1), w_t = (a_2, b_2)] = ((t-s)/12)((a_2 - a_1)^2 + (b_2 - b_1)^2) + (t-s)^2/6$.

Put

$$F_n = \mathbb{E}[F|\mathscr{H}_{2^n}] = \sum_k \frac{w^1_{k/2^n} + w^1_{(k+1)/2^n}}{2}(w^2_{(k+1)/2^n} - w^2_{k/2^n}).$$

Of course $F_n \to F$ almost surely and in L^p, $p \in]1, \infty[$.

(iii) $\mathbb{E}[(F-F_n)^2|\mathscr{H}_{2^n}] = \frac{1}{12}\,1/2^n \sum_{k=0}^{2^n-1} |w((k+1)/2^n) - w(k/2^n)|^2 + \frac{1}{6}1/2^n.$

Then putting V_n^2 for this last expression, we define

$$S_\epsilon = \inf\{n \geq 0 : V_n^2 \leq \epsilon^2\}.$$

By the same argument as in the preceding example, F_{S_ϵ} is an s-functional. We obviously have $\mathbb{E}[(F - F_{S_\epsilon})^2|\mathscr{H}_{2S_\epsilon}] \leq \epsilon^2$ and, taking the expectation,

$$\|F - F_{S_\epsilon}\|_2 \leq \epsilon.$$

D.5.5 Presentation with General Partitions

Rather than using the dyadic rationals, we can instead use all the rational numbers. Let $r : \mathbf{N} \to \mathbf{Q} \cap [0, 1]$ be an enumeration of the rational numbers $p/q \in [0, 1]$ and put

$$X_n \circ \varphi(w) = w(r(n))$$

With this presentation, it is possible to define s-functionals, which depend on the points of a partition of [0, 1] whose "deepth" is random, as before, but whose "thinness" can now vary from place to place in [0, 1] as a function of the values at points already picked. This family includes the discretization for diffusion in space studied by O. Faure [1990]. Of course, many other presentations of Wiener space can be constructed, each with its own family of s-functionals.

D.5.6 Remark

If we consider the program that yields a picture like that at the end of Chapter 3 showing a 2-dimensional Brownian path (which is done here by recursive simulation), we get a map from the Wiener space to the Polish space of piecewise affine lines of the plane. As it is now clear, this is an s-functional.

E SYMBOLIC EXPANSIONS OF SOLUTIONS TO SDE's

In this section, we consider the algebraic features of the stochastic calculus. Our account is purely a symbolic survey, we do not detail the assumptions, which must be stated for these formulas to be valid, and are still partly in the domain of research.

E.1 Ito–Stratonovich Transformation

Definition E.1.1 Let X and Y be continuous semi-martingales. The Stratonovich stochastic integral is defined by

$$\int_0^t X_s \circ dY_s = \int_0^t X_s dY_s + \tfrac{1}{2}\langle X, Y\rangle_t$$

The following chain rule follows immediately from Ito's formula.

Proposition E.1.2 Let $X^1, \ldots, X^n$ be continuous semi-martingales and let F be a C^3-function from $\mathbb{R}^n$ to $\mathbb{R}$. Then

$$F(X_t^1, \ldots, X_t^n) = F(X_0^1, \ldots, X_0^n) + \sum_{i=1}^{n} \int_0^t F'_i(X_s^1, \ldots, X_s^n) \circ dX_s^i.$$

E.2 Hu–Meyer Formulas

We consider a standard (F_t)-$\mathbb{P}$-Brownian motion $B_t = (B_t^1, \ldots, B_t^d)$ with values in $\mathbb{R}^d$. Let E be the space $\mathbb{R}_+ \times \{1, \ldots, d\}$ equipped with the product measure μ of the Lebesgue measure with the counting measure on $\{1, \ldots, d\}$. Let f_n be in $L^2_{sym}(E^n, \mu^{\otimes n})$ the space of symmetric functions from E^n to $\mathbb{R}$ which are in $L^2(E^n, \mu^{\otimes n})$.

We define $I_n(f_n)$ by

$$I_n(f_n) = n! \sum_{(i_1, \ldots, i_n) \in \{1, \ldots, d\}^n} \int_{0 < t_1 < \cdots < t_n} f_n((t_1, i_1), \ldots, (t_n, i_n))\, dB_{t_1}^{i_1} \ldots dB_{t_n}^{i_n}.$$

The set of random variables of the form $I_n(f_n)$ is a closed subspace

of $L^2(\mathbb{P})$, say, Π_n, called nth *Wiener chaos* and we have the following direct sum

$$L^2(\mathbb{P}) = \bigoplus_{n=0}^{\infty} \Pi_n$$

(cf. Bouleau and Hirsch [1991, Proposition II.2.1.2]). Thus every random variable F in L^2 defines a sequence $\mathbf{f} = (f_n)$ such that $F = \sum_{n=0}^{\infty}(1/n!)I_n(f_n)$.

We set

$$I(\mathbf{f}) = \sum_{n=0}^{\infty} \frac{1}{n!} I_n(f_n).$$

The isometry of stochastic integrals then yields

$$\|I(\mathbf{f})\|^2 = \sum_{n=0}^{\infty} \frac{1}{n!} \|f_n\|^2.$$

The norm of the left-hand side is taken in $L^2(\mathbb{P})$, and that of the right-hand side in $L^2_{sym}(E^n, \mu^{\otimes n})$.

A trace operator Tr is defined on a suitable subspace of $L^2_{sym}(E^n, \mu^{\otimes n})$ consisting of sufficiently smooth functions, with values in $L^2_{sym}(E^{n-2}, \mu^{\otimes n-2})$, given by

$$Trf_0 = Trf_1 = 0$$

and, for $n \geq 2$,

$$Tr(f_n)((t_1, i_1), \ldots, (t_{n-2}, i_{n-2})) = \sum_{(i_n, i_{n-1}) \in \{1, \ldots, d\}^2} \delta_{i_n, i_{n-1}} \int_{\mathbb{R}_+} f_n((t_1, i_1), \ldots, (t_{n-2}, i_{n-2})(s, i_{n-1})(s, i_n))\, ds,$$

where $\delta_{i,j}$ is the Kronecker symbol: $\delta_{i,j} = 1$ if $i = j$, $\delta_{i,j} = 0$ otherwise.

The operator Tr can be extended to sequences $\mathbf{f} = (f_n)$ by

$$(Tr(\mathbf{f}))_n = Tr(f_{n+2}) \qquad n \geq 0.$$

Let us return now to the definition of $I_n(f_n)$ and suppose that the functions f_n are sufficiently smooth for the multiple Ito integral to be defined as successive simple integrals of processes which are semi-martingales. We can then replace the Ito integrals by Stratonovich integrals, and define

$$S_n(f_n) = n! \sum_{(i_n, i_{n-1}) \in \{1, \dots, d\}^n} \int_{0<t_1<\cdots<t_n} f_n((t_1, i_1) \cdots (t_n, i_n)) \circ dB_{t_1}^{i_1} \cdots \circ dB_{t_n}^{i_n}.$$

For a sequence $\mathbf{f} = (f_n)$ of such functions, we set

$$S(\mathbf{f}) = I(e^{(1/2)Tr}\mathbf{f}). \tag{E.2.1}$$

In this formula the operator $e^{(1/2)Tr} = \sum_{k=0}^{\infty} (1/k!)(1/2^k)Tr^k$ operates on sequences. Taking for $\mathbf{f}$ the sequence where only f_n differs from zero, gives

$$S_n(f_n) = \sum_{0 \le k \le n/2} \frac{n!}{(n-2k)!k!} \frac{1}{2^k} I_{n-2k}(Tr^k f_n). \tag{E.2.2}$$

Similarly applying (E.2.1) to the sequence $e^{(1/2)Tr}\mathbf{f}$ gives

$$I(\mathbf{f}) = S(e^{-(1/2)Tr}\mathbf{f}) \tag{E.2.3}$$

and

$$I_n(f_n) = \sum_{0 \le k \le n/2} \frac{n!}{(n-2k)!k!} \left(-\frac{1}{2}\right)^k S_{n-2k}(Tr^k f_n). \tag{E.2.4}$$

Formulas (E.2.1) to (E.2.4) are called Hu-Meyer formulas [Hu and Meyer, 1988]. We shall explain these formulas in the simple case when $d = 1$ and $f_n = h^{\otimes n}$ for $h \in L^2(\mathbb{R}_+)$:

$$f_n(t_1, \dots, t_n) = h(t_1) \cdots h(t_n).$$

If as before we set

$$\mathbf{f} = (f_n)_{n \ge 0} = (h^{\otimes n})_{n \ge 0}$$

then, since the Stratonovich integral satisfies the usual chain rule (Proposition E.1.2), we have

$$S(\mathbf{f}) = \exp\left[\int_0^\infty h(s)\,dB_s\right] \tag{E.2.5}$$

and

$$S_n(f_n) = \left(\int_0^\infty h(s)\,dB_s\right)^n. \tag{E.2.6}$$

Moreover, if we consider the martingale

$$M_t = \exp\left[\int_0^t h(s)\,dB_s - \tfrac{1}{2}\int_0^t h(s)^2\,ds\right],$$

then the Ito formula shows that it satisfies the equation

$$M_t = 1 + \int_0^t M_s h(s)\,dB_s.$$

This is also satisfied by the martingale N_t defined by

$$N_t = 1 + \sum_{n\geq 1}\int_{0<t_1<\cdots<t_n<t} h(t_1)\cdots h(t_n)\,dB_{t_1}\cdots dB_{t_n}.$$

It follows [Bouleau and Hirsch, 1991, Lemma II.2.1.2.1] that $M_t = N_t$. Thus we have

$$I(\mathbf{f}) = \exp\left[\int_0^\infty h(s)\,dB_s - \tfrac{1}{2}\int_0^\infty h(s)^2\,ds\right]. \tag{E.2.7}$$

Let $z \in \mathbf{C}$, and expand $\exp\left[z\int h(s)\,dB_s - z^2/2\int h^2(s)\,ds\right]$ as a power series in z. We obtain

$$\sum_{n=0}^{\infty}\frac{1}{n!}\left(z\int h\,dB-\frac{z^2}{2}\int h^2\,ds\right)^n$$
$$=\sum_{m=0}^{\infty}z^m\sum_{0\le k\le m/2}\left(-\frac{1}{2}\right)^k\frac{1}{(m-2k)!k!}\left(\int h\,dB\right)^{m-2k}\left(\int h^2\,ds\right)^k.$$

By (E.2.7) the coefficient of $z^m/m!$ is equal to $I_m(f_m)$. Moreover, by (E.2.6)

$$S_{m-2k}(h^{\otimes m-2k})=\left(\int h\,dB\right)^{m-2k}$$

and

$$S_{m-2k}(Tr^k h^{\otimes m})=\left(\int h\,dB\right)^{m-2k}\left(\int h^2\,ds\right)^k.$$

So we have

$$I_m(f_m)=\sum_{0\le k\le m/2}\left(-\frac{1}{2}\right)^k\frac{1}{(m-2k)!k!}S_{m-2k}(Tr^k f_m),$$

which is (E.2.4), and the other formulas (E.2.3), (E.2.1), and (E.2.2) follow.

E.3 Isobe–Sato Formula

Let $B_t=(B_t^1,\dots,B_t^d)$ be a d-dimensional standard Brownian motion with respect to the filtration $(\mathscr{F}_t)$. We consider a stochastic differential equation in Stratonovich form

$$\text{(E.3.1)}\qquad \begin{cases} dX_t^i(x)=\displaystyle\sum_{\alpha=1}^{d}\sigma_\alpha^i(X_t(x))\circ dB_t^\alpha+\sigma_0^i(X_t(x))\,dt \\ X_0(x)=x, \end{cases}$$

where $X_t(x)=(X_t^1(x),\dots,X_t^n(x))$ takes values in $\mathbb{R}^n$, and the coefficients σ_α^i, σ_0^i $i=1,\dots,n$ are C^∞. Equation (E.3.1) is a rewriting of the Ito equation

$$\text{(E.3.2)} \qquad \begin{cases} dX_t^i(x) = \sum_{\alpha=1}^{d} \sigma_\alpha^i(X_t(x))\, dB_t^\alpha + \bar{\sigma}_0^i(X_t(x))\, dt \\ X_0(x) = x \end{cases}$$

with $\bar{\sigma}_0^i(x) = \sigma_0^i(x) + \frac{1}{2}\sum_{\alpha=1}^{d}\sum_{k=1}^{d}((\partial/\partial x^k)\sigma_\alpha^i(x))\sigma_\alpha^k(x)$.

By an extension of Ito's theorem [Ikeda and Watanabe, 1981, Chapter IV, Theorem 3.1] equation (E.3.2) can be shown to have a unique strong solution (cf. Ikeda and Watanabe [1981, Chapter IV, Definition 1.6] or Rogers and Williams [1987, Chapter V, Section 2], or Revuz and Yor [1991, Chapter IX]). If, in addition, the coefficients of equation (E.3.1) are globally Lipschitz, the solution ([Ikeda and Watanabe, 1981, Chapter IV, Section 6] defines a strong Markov process with generator A given by

$$Af(x) = \frac{1}{2}\sum_{i,j} a_{ij}(x)\frac{\partial^2 f}{\partial x^i \partial x^j}(x) + \sum_i b_i(x)\frac{\partial f}{\partial x_i}(x),$$

where

$$a_{ij}(x) = \sum_\alpha \sigma_\alpha^i(x)\sigma^j{}_\alpha(x) \quad \text{and} \quad b_i(x) = \bar{\sigma}_0^i(x) \quad \text{and} \quad f \in \mathscr{C}_c^2(\mathbb{R}^n).$$

Let A_α be the field of first-order differential operators defined by

$$A_\alpha f(x) = \sum_{i=1}^{n} \sigma_\alpha^i(x)\frac{\partial f}{\partial x^i} \qquad \alpha \geq 1$$

and

$$A_0 f(x) = \sum_{i=1}^{n} \sigma_0^i(x)\frac{\partial f}{\partial x^i}.$$

Then we have the following lemma.

Lemma E.3.3 $\forall f \in \mathscr{C}_c^2(\mathbb{R}^n)$

$$Af = \tfrac{1}{2}\sum_{\alpha=1}^{d} A_\alpha(A_\alpha f) + A_0 f.$$

Proof By (E.3.2) we have

$$\overline{\sigma}_0^i = \sigma_0^i + \sum_{\alpha=1}^{d} A_\alpha \sigma_\alpha^i$$

and the lemma follows. □

We consider a smooth function $h : \mathbb{R}^n \to \mathbb{R}$ and write the Wiener chaos expansion of the random variable $h(X_T(x))$, $T \in \mathbb{R}_+$, as

$$h(X_T(x)) = \sum_{m=0}^{\infty} \sum_{(\alpha_1, \ldots, \alpha_m) \in \{1, \ldots, d\}^m} \int_{0<s_1<\ldots<s_n<T} \cdot F_m(x, T, h; (s_1, \alpha_1), \ldots, (s_m, \alpha_m)) dB_{s_1}^{\alpha_1} \cdots dB_{s_m}^{\alpha_m}.$$

The Isobe–Sato formula yields an expression of the functions F_m in terms of the semi-group (P_t) of the Markov process X and the operators A_α, $\alpha = 1, \ldots, d$:

$$\begin{aligned} &F_m(x, T, h; (s_1, \alpha_1), \ldots, (s_m, \alpha_m)) \\ &\quad = P_{s_1} A_{\alpha_1} P_{s_2 - s_1} A_{\alpha_2} \cdots P_{s_m - s_{m-1}} A_{\alpha_m} P_{T - s_m} h(x). \end{aligned} \tag{E.3.4}$$

As before, we explain this formula in the case $d = n = 1$. We follow the method sketched in Léandre and Meyer [1989].

Consider first any random variable $f \in L^2$ with chaos expansion

$$f = \sum_{m-0}^{\infty} \int_{s_1 < \cdots < s_m} f_m(s_1, \ldots, s_m)\, dB_{s_1} \cdots dB_{s_m}.$$

If g_1 is continuous with compact support on $\mathbb{R}_+$, we have

$$\mathbb{E}\left(f \int_0^\infty g_1(t)\, dB_t \right) = \int_0^\infty f_1(t) g_1(t)\, dt \tag{E.3.5}$$

and, similarly,

$$\mathbb{E}[f\int_{t_1<\cdots<t_m} g_m(t_1,\ldots,t_m)\,dB_{t_1}\cdots dB_{t_m}]$$

$$= \int_{t_1<\cdots<t_m} f_m(t_1,\ldots,t_m)g_m(t_1,\ldots,t_m)\,dt_1\cdots dt_m \tag{E.3.6}$$

for g_m continuous with compact support in the sector $\Delta_m = \{(t_1,\ldots,t_m) : 0\le t_1<\ldots<t_m\}$. Thus the map

$$g_m \rightsquigarrow \mathbb{E}\left[f\int_{t_1<\cdots<t_m} g_m(t_1,\ldots,t_m)\,dB_{t_1}\cdots dB_{t_m}\right]$$

defines a bounded measure on Δ_m, absolutely continuous with respect to $dt_1\cdots dt_m$ and with density $f_m(t_1,\ldots,t_m)$. It is convenient to write this property in the symbolic differential form :

$$\mathbb{E}[f.dB_{t_1}\cdots dB_{t_m}] = f_m(t_1,\ldots,t_m)\,dt_1\cdots dt_m. \tag{E.3.7}$$

Let us return now to our diffusion process ($d = n = 1$) and compute using (E.3.7) the chaos expansion of $h(X_T(x))$.

Beginning with the first chaos ($m = 1$), we have

$$\begin{aligned} F_1(x,T,h,t_1)dt_1 &= \mathbb{E}[h(X_T(x)).dB_{t_1}] \\ &= \mathbb{E}[\mathbb{E}[h(X_T(x))|\mathcal{F}_{t_1}].dB_{t_1}] \\ &= \mathbb{E}[P_{T-t_1}h(X_{t_1}(x)).dB_{t_1}]. \end{aligned}$$

Moreover, Ito's formula gives

$$\begin{aligned} P_{T-t_1}h(X_{t_1}(x)) = P_{T-t_1}h(x) &+ \int_0^{t_1}(P_{T-t_1}h)'(X_s(x))\sigma(X_s(x))\,dB_s \\ &+ \int_0^{t_1}(P_{T-t_1}h)'(X_s(x))b(X_s(x))\,ds \\ &+ \tfrac{1}{2}\int_0^{t_1}(P_{T-t_1}h)''(X_s(x))\sigma^2(X_s(x))\,ds. \end{aligned}$$

We use this relation to compute $\mathbb{E}[P_{T-t_1}h(X_{t_1}(x)).dB_{t_1}]$. We get

$$\mathbb{E}[P_{T-t_1}h(x).dB_{t_1}] = 0$$

$$\mathbb{E}\left[\int_0^{t_1}(P_{T-t_1}h)'(X_s(x))(b+\sigma^2)(X_s(x))ds.dB_{t_1}\right] = 0.$$

Thus only

$$\begin{aligned}\mathbb{E}[P_{T-t_1}h(X_{t_1}(x)).dB_{t_1}] &= \mathbb{E}\left[\int_0^{t_1}(P_{T-t_1}h)'(X_s(x))\sigma(X_s(x))\,dB_s.dB_{t_1}\right]\\ &= \mathbb{E}[(P_{T-t_1}h)'(X_{t_1}(x))\sigma(X_{t_1}(x))dt_1]\\ &= P_{t_1}\sigma(P_{T-t_1}h)'(x)dt_1.\end{aligned}$$

remains.

And finally,

$$F_1(x,T,h,t_1) = P_{t_s}AP_{T-t_1}h(x).$$

For general m, the computation is similar. We evaluate $\mathbb{E}[h(X_T(x))\,dB_{t_1}\cdot\cdots dB_{t_m}]$ for $t_1 < \cdots < t_m < T$ using the facts that $dB_{t_{m-1}}$ is $\mathscr{F}_{t_m}$-measurable, $dB_{t_{m-2}}$ is $\mathscr{F}_{t_{m-1}}$-measurable, and so on.

E.4 Expansion of the Solution of a SDE as a Series of Iterated Stochastic Integrals

For notational convenience we will put

$$B_t^0 = t.$$

For $J = (j_1,\ldots,j_m) \in \{0,\ldots,d\}^m$, we define the integrals

$$S_t^J = \int\cdots\int_{0<t_1<\cdots<t_m<t} \circ dB_{t_1}^{j_1}\cdots\circ dB_{t_m}^{j_m}$$

and

$$I_t^J = \int\cdots\int_{0<t_1<\cdots<t_m<t} dB_{t_1}^{j_1}\cdots dB_{t_m}^{j_m}.$$

Recall the Stratonovich equation (E.3.1)

$$dX_t = \sum_{\alpha=0}^{d} \sigma_\alpha(X_t) \circ dB_t^\alpha \qquad X_0 = x \in \mathbb{R}^n, \tag{E.4.1}$$

where we suppose now that the coefficients can be expanded as power series in a ball with center $x = (x^i)$. Our aim is to represent the solution X_t as a series of iterated stochastic integrals. For this we consider the auxiliary equation

$$X_t^\epsilon = x + \epsilon \sum_{\alpha=1}^{d} \int_0^t \sigma_\alpha(X_s^\epsilon) \circ dB_s^\alpha + \epsilon^2 \int_0^t \sigma_0(X_s^\epsilon)\, ds. \tag{E.4.2}$$

Let us write the Taylor expansions at the point x

$$\sigma_\alpha(y) = \sum_{\nu \in \mathbb{N}^n} a_\nu^\alpha (y - x)^\nu \qquad \alpha = 0, 1, \ldots, d$$

where, as usual, $y^\nu = (y^1)^{\nu_1} \cdots (y^n)^{\nu_n}$ for $\nu = (\nu_1, \ldots, \nu_n) \in \mathbf{N}^n$ and $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$. Here the coefficients a_ν^α take values in $\mathbb{R}^n$.

Let us suppose that the process X_t^ϵ can be expanded as a power series in ϵ with coefficients that are processes Y_t^i with values in $\mathbb{R}^n$:

$$X_t^\epsilon = \sum_{i \geq 0} \epsilon^i Y_t^i \quad \text{with} \quad Y_t^0 = x. \tag{E.4.3}$$

In differential form, the equation (E.4.2) becomes

$$\begin{aligned} \sum_{i \geq 0} \epsilon^i dY_t^i = {} & \epsilon \sum_{\nu \in \mathbb{N}^n} \sum_{\alpha=1}^{d} a_\nu^\alpha \left(\sum_{i \geq 1} \epsilon^i Y_t^i \right)^\nu \circ dB_t^\alpha \\ & + \epsilon^2 \sum_{\nu \in \mathbb{N}^n} a_\nu^0 \left(\sum_{i \geq 1} \epsilon^i Y_t^i \right)^\nu dt. \end{aligned} \tag{E.4.4}$$

Now we identify the coefficients of ϵ^i, $i = 1, 2, \ldots$. We obtain

$$
\begin{aligned}
i = 0 \qquad & dY_t^0 = 0 \\
i = 1 \qquad & dY_t^1 = \sum_{\alpha=1}^{d} \sigma_\alpha(x) \circ dB_t^\alpha \\
i = 2 \qquad & dY_t^2 = \sum_{\alpha=1}^{d} \sum_{|\nu|=1} a_\nu^\alpha (Y_t^1)^\nu \circ dB_t^\alpha + \sigma_0(x)\, dt \\
i = 3 \qquad & dY_t^3 = \sum_{\alpha=1}^{d} \left(\sum_{|\nu|=1} a_\nu^\alpha (Y_t^2)^\nu + \sum_{|\nu|=2} a_\nu^\alpha (Y_t^1)^\nu \right) \circ dB_t^\alpha \\
& \qquad + \sum_{|\nu|=1} a_\nu^0 (Y_t^1)^\nu \, dt,
\end{aligned}
$$

and so on.

More precisely, if $b_1, \ldots, b_m$ are points in $\mathbb{R}^n$, we consider the polynomial $P(z) = b_1 z + \cdots + b_m z^m$ for $z \in \mathbb{R}$. Let $C_\alpha^k(b_1, \ldots, b_m)$ be the coefficient of z^k in the formal series

$$
\sigma_\alpha(x + P(z)) = \sum_{\nu \in \mathbb{N}^n} a_\nu^\alpha (b_1 z + \cdots + b_m z^m)^\nu .
$$

Then our identification can be written

$$
\text{(E.4.5)} \quad \begin{cases}
dY_t^1 = \displaystyle\sum_{\alpha=1}^{d} \sigma_\alpha(x) \circ dB_t^\alpha \\
dY_t^2 = \displaystyle\sum_{\alpha=1}^{d} C_\alpha^1(Y_t^1) \circ dB_t^\alpha + \sigma_0(x)\, dt \\
\qquad \vdots \\
dY_t^m = \displaystyle\sum_{\alpha=1}^{d} C_\alpha^{m-1}(Y_t^1, \ldots, Y_t^{m-1}) \circ dB_t^\alpha \\
\qquad\quad + C_0^{m-2}(Y_t^1, \ldots, Y_t^{m-2})\, dt.
\end{cases}
$$

We note that the equation for Y_t^m depends only on $Y_t^1, \ldots, Y_t^{m-1}$. Thus the system of equations which gives Y_t^m can be solved by successive stochastic quadratures, and so Y_t^m appears as a finite sum of iterated stochastic integrals.

For $J = (j_1, \ldots, j_m)$ we put

$$\|J\| = m + card\{i : j_i = 0\}$$

(which amounts to count stochastic integrals for 1 and dt-integrals for 2). Then we have the following proposition:

Proposition E.4.6 Let $(Y_t^m)^j$ be the j-th component of the process Y_t^m. We have

$$(Y_t^m)^j = \sum_{\|J\|=m} P_J^j(a_\nu^\alpha) S_t^J,$$

where the P_J^j are universal polynomials with rational coefficients in the coefficients a_ν^α of the σ_α.

Proof We shall only sketch the argument. We call $\|J\|$ the order of the integral S_t^J.

By successive applications of the chain rule (which is Ito's formula for Stratonovich integrals) we see that a product $S_t^{J_1} S_t^{J_2}$ can be written as a finite sum of stochastic integrals of order $\|J_1\| + \|J_2\|$. An induction argument on i then shows that, by the identification (E.4.5), Y_t^i is a finite sum of stochastic integrals of order i.

(In other words, in the series $X_t = \sum \epsilon^i Y_t^i$ which represents X, the degree of ϵ is the order of Y_t^i). □

Corollary E.4.7 For every $s > 0$, the processes

$$(\sqrt{s} Y_t^1, \ldots, s^{m/2} Y_t^m) \quad \text{and} \quad (Y_{st}^1, \ldots, Y_{st}^m)_{t \geq 0}$$

have the same law.

Proof This is a consequence of the scaling property of the Brownian motion

$$(B_{st})_{t \geq 0} \sim (\sqrt{s} B_t)_{t \geq 0}$$

and the previous representation. □

E.4.8 Example
Let us consider the equation

$$dX_t = X_t \circ dB_t, X_0 = x,$$

which, in Ito's form, gives

$$dX_t = X_t dB_t + \tfrac{1}{2} X_t dt.$$

The Stratonovich form, which follows the ordinary rules for change of variable, immediately gives the solution

$$X_t = xe^{B_t}.$$

The formal expansion

$$X_t = x(1 + \cdots + \int_{0<t_1<\cdots<t_n<t} \circ dB_{t_1} \cdots \circ dB_{t_n})$$
$$= x \sum_{n=0}^{\infty} \frac{1}{n!} B_t^n$$

converges for every t.

E.4.9 Example
Consider the equation

$$\begin{cases} dX_t = X^q \circ dB_t \\ X_0 = 1. \end{cases} \qquad q \geq 2$$

We obtain

$$X_t = \left(\frac{1}{1 - (q-1)B_t} \right)^{1/(q-1)},$$

which can be expanded as a power series of B_t by composing the expansions of $1/1 - x$ and $(1 + x)^{\alpha}$. For $q = 2$, the expansion is

$$1 + B_t + \cdots + B_t^k + \cdots .$$

It diverges, for fixed t, with positive probability, as does $\sum \|B_t^k\|_2$. But the series converges for t strictly smaller than the stopping time $T = \inf\{t = |B_t| = 1\}$.

F APPLICATION OF THE SHIFT METHOD TO MULTIPLE WIENER INTEGRALS AND TO SOLUTIONS OF SDEs

The shift method was given in a general setting in Chapter 3, Section A.4. It is based on Birkhoff's pointwise ergodic theorem. This method is especially interesting because its implementation enables us to store intermediate computations. We investigated the rate of convergence of this method in Chapter 3, Section A.5. Not every function in L^2 satisfies a law of the iterated logarithm, and in Chapter 3, Section A.5.3 we introduced the subset of L^2 called the Gordin class where an L.I.L. holds.

Our purpose here is to investigate the application of this method to computing the expectations of functions of multiple Wiener integrals or solutions of stochastic differential equations of the Ito type, and to examine whether or not these functionals belong to the Gordin class.

We use the notation of Chapter 3, Section A.5.3:

$(E, \mathscr{E}, \mu)$ is a probability space and $(\Omega, \mathscr{A}, \mathbb{P}) = (E^{\mathbf{Z}}, \mathscr{E}^{\otimes \mathbf{Z}}, \mu^{\otimes \mathbf{Z}})$.

The coordinate functions from Ω to E are denoted by X_n. We define the ergodic automorphism τ on $(\Omega, \mathscr{A}, \mathbb{P})$ by

$$X_n \circ \tau = X_{n-1} \qquad \forall n \in \mathbf{Z}.$$

We call τ the shift to the right. We put

$$\mathscr{F}_m^n = \sigma(X_m, X_{m+1}, \ldots, X_n) \qquad \text{for} \quad m \leq n \in \mathbf{Z}$$
$$\mathscr{F}_{-\infty}^n = \sigma(X_k, k \leq n)$$
$$\mathscr{F}_m^{+\infty} = \sigma(X_k, k \geq m)$$
$$\mathscr{F}_{-\infty}^{+\infty} = \sigma(X_k, k \in \mathbf{Z}).$$

Let T be the operator on $L^2(\mathscr{F}_0^\infty, \mathbb{P})$ defined by

$$Tf = \mathbb{E}[f | \mathscr{F}_1^\infty] \circ \tau.$$

We denote by $\mathcal{G}$ the Gordin class (cf. Chapter 3, Section A.6).

To apply the results of Chapter 3, Section A.6, we must write the Wiener space as a product space on which the shift operates. We can do this in several ways that give different results. We first study a factorization in which the shift appears as a scaling transform on the Wiener space.

F.1 First Factorization of the Wiener Space

Consider the space $W = \{f \in \mathcal{C}([0,1], \mathbb{R}^d), f(0) = 0\}$ equipped with its Borel σ-field $\mathcal{B}$ and with the Wiener measure m.

On the space $(\Omega, \mathcal{A}, \mathbb{P}) = \prod_{n=-\infty}^{+\infty}(W_n, \mathcal{B}_n, m_n)$, where $(W_n, \mathcal{B}_n, m_n)$ are copies of $(W, \mathcal{B}, m)$, we define a Brownian motion $(B_t)_{t\in\mathbb{R}_+}$ as follows. Let X_n be the coordinate map from Ω into W_n and for $t \in]1/2^{k+1}, 1/2^k]$ and $k \in \mathbf{Z}$ we put

$$B_t = \left(\sum_{n=k+1}^{\infty} \frac{X_n(1)}{2^{(n+1)/2}} \right) + \frac{X_k((t - 1/2^{k+1})/(1/2^{k+1}))}{2^{(k+1)/2}}.$$

The process thus defined is Gaussian centered with independent increments, and tends to zero as t goes to zero. Its covariance is easily computed to be $s \wedge t$ times the identity matrix ; it is therefore a standard $\mathbb{R}^d$-valued Brownian motion.

The transform τ defined on Ω by

$$X_n \circ \tau = X_{n-1} \qquad n \in \mathbf{Z}$$

is a scaling:

$$B_t \circ \tau = \frac{1}{\sqrt{2}} B_{2t}.$$

The results of Chapter 3, Section A.6 apply with

$$\mathcal{F}_0^\infty = \sigma(B_s, s \le 1)$$
$$\mathcal{F}_1^\infty = \sigma(B_s, s \le \tfrac{1}{2})$$
$$\mathcal{F}_0^k = \sigma(B_s - B_{1/2^{k+1}}, s \in]1/2^{k+1}, 1]).$$

F.2 Functions of Solutions of Lipschitz SDEs

We consider maps

$$\sigma : \mathbb{R}^m \times \mathbb{R}_+ \to \mathbb{R}^{m\times d} \quad \text{and} \quad b : \mathbb{R}^m \times \mathbb{R}_+ \to \mathbb{R}^m$$

that satisfy the Lipschitz hypotheses: $\exists C > 0$ such that $\forall s \in [0,1] \forall x, y \in \mathbb{R}^m$

$$|\sigma(x,s)| + |b(x,s)| \le C(1+|x|)$$

$$|\sigma(x,s) - \sigma(y,s)| + |b(x,s) - b(y,s)| \le C|x-y|,$$

where $|.|$ is one of the equivalent norms on the Euclidean spaces.

Let X_t^x be the solution of the SDE

$$X_t^x = x + \int_0^t \sigma(X_s^x, s)\, dB_s + \int_0^t b(X_s^x, s)\, ds \qquad x \in \mathbb{R}^m.$$

Proposition F.2.1 Let $f = h(X_t^x)$ for $t \le 1$ with $h : \mathbb{R}^m \to \mathbb{R}$ Hölderian with exponent $\lambda \in]0,1]$. Then $f \in \mathscr{G}$.

Proof Let A be the Hölder constant of h:

$$|h(x) - h(y)| \le A|x-y|^\lambda,$$

and let $(P_t)_{t\le 0}$ be the semi-group of the Markov diffusion process associated with the solution X_t^x. By classical estimates [Kunita, 1982, Chapter 2], we have

$$\begin{aligned} |P_u h(x) - P_u h(y)| &= |\mathbb{E}h(X_u^x) - \mathbb{E}h(X_u^y)| \le A\mathbb{E}|X_u^x - X_u^y|^\lambda \\ &\le K|x-y|^\lambda \qquad \forall u \in [0,1], \end{aligned}$$

for some constant K depending on the dimensions m and d and on the constants C and A.

If φ is Hölder with exponent λ, we have

$$\begin{aligned}\mathrm{Var}\,[\varphi(X_s^x)] &= \mathbb{E}|\varphi(X_s^x) - \mathbb{E}\varphi(X_s^x)|^2 \le \mathbb{E}|\varphi(X_s^x) - \varphi(\mathbb{E}(X_s^x))|^2 \\ &\le C_1\mathbb{E}|X_s^x - \mathbb{E}X_s^x|^{2\lambda} \le C_2(1 + |x|^{2\lambda})s^{\lambda}\end{aligned}$$

(see Kunita [1982, Theorem 2.1]).

Now, we put $\mathscr{B}_t = \sigma(B_s, s \le t)$ and remark that

$$T^n f = \mathbb{E}[h(X_t^x)|\mathscr{B}_{1/2^n}] \circ \tau^n = P_{t-(1/2^n)}h(X_{1/2^n}^x) \circ \tau^n.$$

Hence by the preceding estimates we get

$$\begin{aligned}\|T^n(f - \mathbb{E}f)\|^2 &= \mathrm{Var}\,[T^n f] = \mathrm{Var}\,[P_{t-(1/2^n)}h(X_{1/2^n}^x)] \\ &\le c(1 + |x|^{2\lambda})\frac{1}{2^{n\lambda}}\end{aligned}$$

and the series $\sum \|T^n(f - \mathbb{E}f)\|$ converges geometrically, which implies the result by Lemma A.6.4 of Chapter 3. □

Proposition F.2.2 Let $\mu(ds, dx)$ be a bounded measure on $[0, 1]\times\mathbb{R}^m$ such that

$$\int_{[0,1]\times\mathbb{R}^m} (1 + |x|^{\lambda})|\mu|(ds, dx) < +\infty$$

with $\lambda \in]0, 1]$, and let g be a Hölderian function of exponent λ. Then the functional

$$f = \int_{[0,1]\times\mathbb{R}^m} g(X_s^x)\mu(ds, dx)$$

belongs to $\mathscr{S}$.

Proof We have

$$T^n f = \int_0^{1/2^n} \int_{\mathbb{R}^m} g(X_s^x)\mu(ds, dx) \circ \tau^n + \int_{1/2^n}^1 \int_{\mathbb{R}^m} P_{s-(1/2^n)} g(X_{1/2^n})\mu\,(ds, dx) \circ \tau^n$$

and hence

$$\|T^n(f - \mathbb{E}f)\| \leq \int_0^{1/2^n} \int_{\mathbb{R}^m} \|g(X_s^x) - g(\mathbb{E}X_s^x)\| |d\mu| + \int_{1/2^n}^1 \int_{\mathbb{R}^m} (\mathrm{Var}\,[P_{s-(1/2^n)} g(X_{1/2^n})])^{1/2} |\, d\mu|.$$

Therefore, by the estimates used in the preceding proof,

$$\|T^n(f - \mathbb{E}f)\| \leq \int_0^{1/2^n} \int_{\mathbb{R}^m} A(1 + |x|^\lambda) s^{\lambda/2} |\mu|\,(ds, dx) + \int_{1/2^n}^1 \int_{\mathbb{R}^m} B(1 + |x|)^\lambda \frac{1}{2^{n\lambda/2}}\, |\mu|\,(ds, dx).$$

By our hypotheses the second term is bounded by $C(1/2^{n\lambda/2})$. For the first term, we remark that

$$\sum_{n=0}^{\infty} 1_{[0,1/2^n]}(s) s^{\lambda/2}$$

is bounded on $s \in [0, 1]$, from which it follows that

$$\sum \|T^n(f - \mathbb{E}f)\| < +\infty.$$

The result follows by Lemma A.6.4 of Chapter 3. □

F.3 Functions of Multiple Wiener Integrals

The setting is the same as in Sections F.1 and F.2.

Let

$$F = \int_{0<t_1<\cdots<t_m<1} h(t_1,\ldots,t_m)\, dB_{t_1}^{i_1} dB_{t_2}^{i_2} \cdots dB_{t_m}^{i_m},$$

where $i_k \in \{1,2,\ldots,d\}$ for $k = 1,\ldots,m$ with

$$\int_{0<t_1<\cdots<t_m<1} h^2(t_1,\ldots,t_m)\, dt_1 dt_2 \cdots dt_m < +\infty.$$

It is easily seen that

$$T^n F = \int_{0<t_1<\cdots<t_m<1} \frac{1}{2^{nm/2}} h\left(\frac{t_1}{2^n},\ldots,\frac{t_m}{2^n}\right) dB_{t_1}^{i_1} dB_{t_2}^{i_2} \cdots dB_{t_m}^{i_m}$$

Therefore F belongs to the Gordin class if and only if

$$\sup_N \int_{0<t_1<\cdots<t_m<1} \left(\sum_{n=0}^{N} \frac{1}{2^{nm/2}} h\left(\frac{t_1}{2^n},\ldots,\frac{t_m}{2^n}\right)\right)^2 dt_1 dt_2 \cdots dt_m < +\infty.$$

F.3.1 Example
Take $m = 1$, $h(x) = 1/x^\alpha$ and $\alpha < \frac{1}{2}$. It is easily seen that

$$F = \int_0^1 \frac{1}{t^\alpha} dB_t \in \mathcal{G} \qquad \forall \alpha < \frac{1}{2}.$$

F.3.2 Example
Let us take

$$h(x) = \frac{1}{\sqrt{x}(-\log x)^\beta}, \qquad \text{with} \quad \beta > \frac{1}{2}.$$

Then

$$F = \int_0^{1/2} \frac{1}{\sqrt{t}(-\log t)^\beta} dB_t$$

is in the Gordin class if $\beta > 1$, but F is not in $\mathcal{G}$ if $\beta \in]\frac{1}{2}, 1]$ although in that case $h \in L^2[0,1]$.

F.3.3 Example

If we take

$$h(x) = \frac{1}{\sqrt{x}} \frac{\sin(\pi \log_2 x)}{\log x},$$

then the functional $F = \int_0^{1/2} h(t)\, dB_t$ gives an example of a functional in $\mathcal{G}$ such that

$$\sum_n \|T^n F\| = +\infty$$

and

$$\int_0^{1/2} |h(t)| dB_t$$

is not in $\mathcal{G}$.

F.3.4 Example

We consider a real Brownian motion ($d = 1$), and a square integrable function F with the following Wiener chaos expansion:

$$\begin{aligned} F &= F_0 + \sum_m F_m \\ &= F_0 + \sum_{m=1}^{\infty} \int_{0<t_1<\cdots<t_m<1} h_m(t_1, \ldots, t_m)\, dB_{t_1} dB_{t_2} \cdots dB_{t_m}. \end{aligned}$$

Suppose that $|h_m(t_1, \ldots, t_m)| \le a_m / t_1^{\alpha_1^m} \cdots t_m^{\alpha_m^m}$, with $\alpha_i^m < \frac{1}{2} \forall i = 1, \ldots, m$. Since the chaos are invariant by T, we get

$$\begin{aligned} \left\| \sum_{n=0}^{N} T^n (F - F_0) \right\|^2 &= \sum_{m=1}^{\infty} \left\| \sum_{n=0}^{N} T^n F_m \right\|^2 \\ &\le \sum_{m=1}^{\infty} \frac{a_m^2}{[1 - 2^{\alpha_1^m + \cdots + \alpha_m^m - m/2}]^2 \prod_{i=1}^{m} \left(i - 2 \sum_{k=1}^{i} \alpha_k^m \right)} \end{aligned}$$

so that, if all α_i^m are equal to $\alpha < \frac{1}{2}$, $F \in \mathcal{G}$ whenever the series

$$\sum_{m=1}^{\infty} \frac{a_m^2}{m!(1-2\alpha)^m}$$

converges.

F.4 Other Factorizations of Wiener Space

Let $(\chi_n(t))_{n\geq 0}$ be an orthonormal basis in $L^2[0,1]$, and let $\varphi_n(t) = \int_0^t \chi_n(s)\,ds$. Let $(g_n)_{n\geq 0}$ be a sequence of independent standard Gaussian variables constructed as the coordinates of $(\Omega, \mathscr{A}, \mathbb{P}) = (\mathbb{R}^{\mathbb{N}}, \mathscr{B}(\mathbb{R}^{\mathbb{N}}), N(0,1)^{\otimes \mathbb{N}})$. The series

$$\sum_n g_n \varphi_n \tag{F.4.1}$$

converges in $\mathscr{C}([0,1])$ a.s. and in $L^p((\Omega, \mathscr{A}, \mathbb{P}), \mathscr{C}([0,1]))$, $p \in [1, \infty[$, and its sum is a Brownian motion under $\mathbb{P}$.

Indeed, if we put $\tilde{\chi}_n(\omega) = \int_0^1 \chi_n(s)\,dB_s$ on the Wiener space and $\mathscr{F}_n = \sigma(\tilde{\chi}_k, k \leq n)$, then, writing B for the identity map from $\mathscr{C}[0,1]$ into itself, we obtain

$$\mathbb{E}[B|\mathscr{F}_n] = \sum_{k=0}^{n} \tilde{\chi}_k \varphi_k, \tag{F.4.2}$$

which can be seen by applying a continuous linear functional μ on $\mathscr{C}[0,1]$ to both sides of (F.4.2) and remarking that $(\mu, \tilde{\chi}_0, \ldots, \tilde{\chi}_n)$ is a Gaussian array. By the convergence properties of vector martingales [Neveu, 1972] we therefore have

$$B = \sum_{k=0}^{\infty} \tilde{\chi}_k \varphi_k \tag{F.4.3}$$

a.s. and in L^p. Since the family of partial sums of the series (F.4.1) has the same law as the sums of (F.4.2), the assertion is proved. □

Representing the Brownian motion as

$$B = \sum_{k=0}^{\infty} \tilde{\chi}_k \varphi_k$$

allows us to define the shift, and the associated Gordin class clearly depends on which basis (χ_n) we chose.

The case of Haar functions is particularly interesting. Let us put $\chi = 1_{[0,1/2[} - 1_{[1/2,1[}$ and

$$\chi_{m,k}(t) = 2^{m/2}\chi(2^m t - k)$$
$$\varphi_{m,k}(t) = \int_0^t \chi_{m,k}(s)\,ds$$

for $t \in \mathbb{R}_+$, $m \in \mathbf{Z}$, and $k \in \mathbf{N}$.

The functions $(\chi_{m,k})_{m\in\mathbf{Z},k\in\mathbf{N}}$ are an orthonormal basis for $L^2(\mathbb{R}_+)$. If $g_{m,k}$ are standard independent Gaussian variables, the Brownian motion can be represented by

$$B(\omega, t) = \sum_{m=-\infty}^{+\infty}\left(\sum_{k=0}^{\infty}\varphi_{m,k}(t)g_{m,k}(\omega)\right),$$

and the scaling studied in Sections F.1, F.2, and F.3 is the mapping that transforms the sequence

$$(g_{m,k}(\omega))_{m,k}$$

into the sequence

$$(g_{m-1,k}(\omega))_{m,k}.$$

The space generated by the functions $(\chi_{m,k})_{m\geq 0,0\leq k<2^m}$ is the subspace of $L^2[0,1]$ orthogonal to the constant functions, and the process

$$Z_t = \sum_{m=0}^{\infty}\sum_{k=0}^{2^m-1}\varphi_{m,k}(t)g_{m,k} \tag{F.4.4}$$

is a standard Brownian bridge vanishing at zero and one. The representation (F.4.4) is unique and converges in $\mathscr{C}_{00}[0,1], = \{f \in \mathscr{C}[0,1], f(0) = f(1) = 0\}$. The functions $\varphi_{m,k}$ constitute a Schauder basis for this space. If $f \in \mathscr{C}_{00}[0,1]$ with

$$f(t) = \sum_{m=0}^{\infty}\sum_{k=0}^{2^m-1}\varphi_{m,k}(t)a_{m,k}(f), \tag{F.4.5}$$

then

$$a_{m,k}(f) = \left[2f\left(\frac{k}{2^m} + \frac{1}{2^m+1}\right) - f\left(\frac{k}{2^m}\right) - f\left(\frac{k+1}{2^m}\right)\right]2^{m/2}.$$

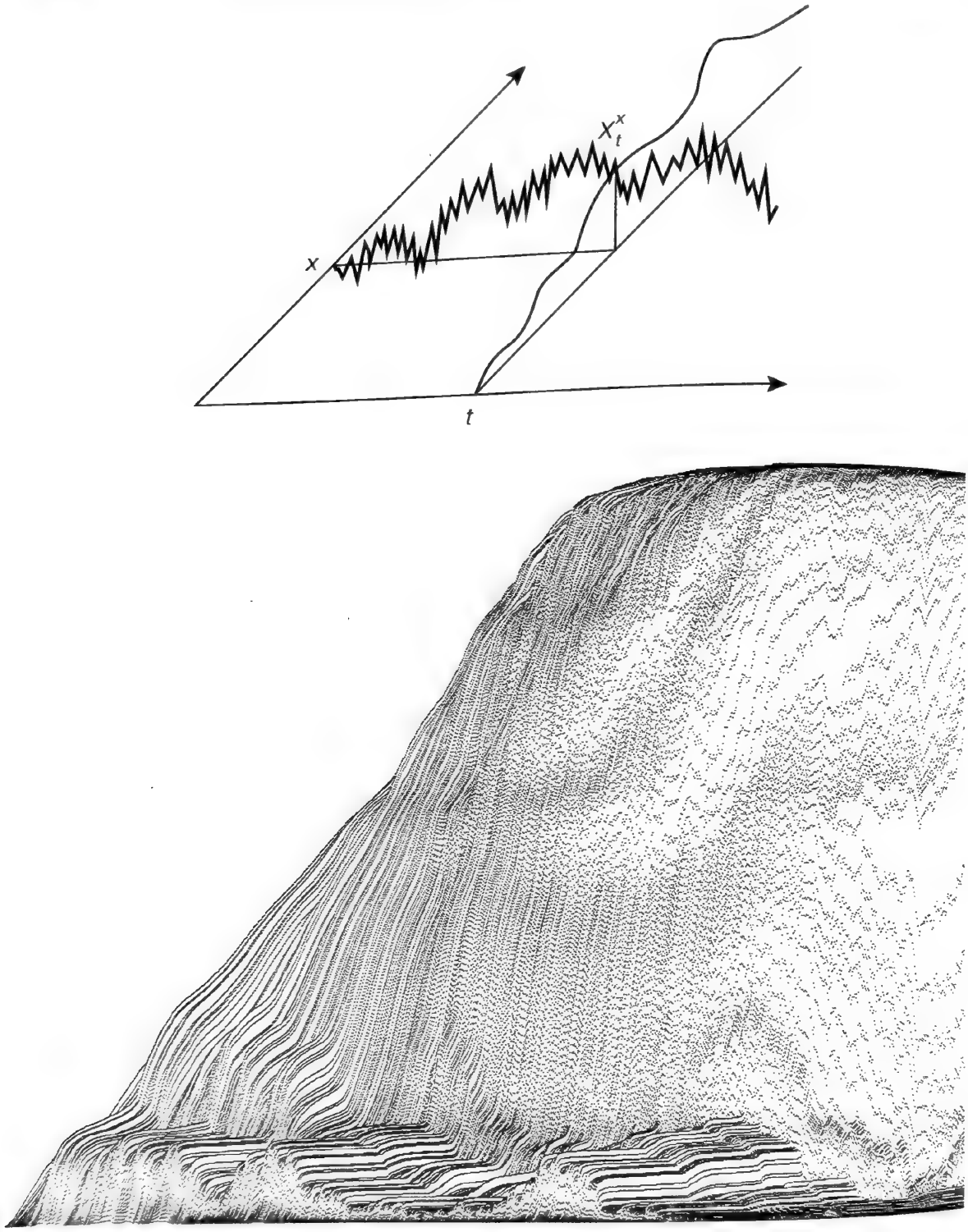


Figure 5.1 The flow of a Lipschitzian stochastic differential equation is the random map $(t,x) \to X_t^x(\omega)$ where X_t^x is the solution of

$$X_t^x = x + \int_0^t \sigma(X_s^x)\, dB_s + \int_0^t b(X_s^x)\, ds.$$

This map is regular with respect to x (smooth if the coefficients σ and b are smooth) and irregular (as a diffusion path) with respect to t. The picture shows the graph of the map $(t,x) \to X_t^x(\omega)$ with hidden parts.

The Banach space of Hölderian functions with exponent $\alpha \in]0, 1[$, which belong to $\mathscr{C}_{00}[0, 1]$, can be interpreted in terms of spaces l^∞ and c_0 on the sequences $(2^{m\alpha} a_{m,k})_{m,k}$ [Ciesielki, 1960].

Approaching a continuous function by a partial sum (F.4.5) is in practice convenient, and if we change the notation by putting $a_{2^m+k} = a_{m,k}$ for $m \geq 0$ and $k = 0, \ldots, 2^m - 1$ the simple shift on the a_n, that is, the transform

$$F(a_1, \ldots, a_n, \ldots) \to F \circ \tau = F(a_0, a_1, \ldots, a_{n+1}, \ldots)$$

(which does not correspond to a scaling) is quite thrifty in random drawings.

Notes

CHAPTER 1

The preliminary material contained in Chapter 1 is covered by a number of works, among which we mention Breiman [1968], Feller [1966], Shiryayev [1984], Bouleau [1988], Dacunha-Castelle and Duflo [1983].

The main results of Chapters 2, 3, and 5 are presented in a simplified form in the article by Bouleau [1993].

CHAPTER 2

It seemed to us that it would be of interest to gather in a single chapter all the material relevant to Monte Carlo and quasi-Monte Carlo methods; this was partly because both methods rely to a great extent on the same mathematical concepts, and partly because they are often in practice interchangeable. A discussion of their relative advantages is given in Section C.5.

It seems likely that future work will produce pseudo-random sequences for which certain asymptotic properties have been *proved* in the same way as for uniformly distributed sequences, with properties other than uniform distribution (asymptotic independence, etc.). From the point of view of the numerical analyst, it will be possible to choose from a variety of approaches according to various criteria: precision, speed of calculation, provability of results, generality or specificity of method, and so on.

Section A.1. For more detailed investigation, a proof of Prokhorov's theorem (A.1.6), and applications of narrow convergence to random processes, see Billingsley [1968] and Shiryayev [1984, Chapter 3].

Section A.2. An explanation of μ-Riemann integrable functions is given in Schwartz [1970] and Bourbaki [1965 (2), Chapter IV, §12].

Section A.3. The fundamental theorem of simulation is standard (see Bouleau [1986]); a more general theorem concerning random processes and their convergence is given in Skorohod [1965].

Section A.4. See Breiman [1968] for a treatment of distribution functions.

Section A.5. The main reference for uniform distribution is still Kuipers and Niederreiter [1974]; see also Niederreiter [1978] and Rauzy [1976] (in particular, for other applications of Weyl's lemma).

Section B.1. For a discussion of the notion of random sequence see Dellacherie [1978], Knuth [1981], and Martin-Löf [1966]. The terminology "Borel sequence" for a sequence that is i.i.d. with uniform law on the cube is introduced here with reference to the works of Emile Borel [1958]. There is a large literature on simulation algorithms—see Devroye [1986], Bouleau [1986], Rubinstein [1981, 1986], Ripley [1987], Knuth [1981]—and we refer to this for techniques for reducing variance, which, as mentioned in the introduction, we do not deal with here. There have been some recent improvements in these techniques, in particular for problems of particle transport in physics, where large deviation methods allow optimal reductions in variance for certain types of event; see Buckley [1990], Hahn and Jeruchim [1987], Cottrell, Fort, and Malgouyres [1983], Yarmolik and Demidenko [1988], Dagpunar [1987].

Sections B2, B3, and B4. The theorem of Glivenko–Cantelli is given, for instance, in Dacunha-Castelle and Duflo [1982] and Bouleau [1986]. An accessible treatment of the generalizations to functions of the central limit theorem, usually known collectively as the principle of invariance, is given in Breiman [1968]; see also the survey of Gaenssler and Stute [1979]. The intuitive meaning of the most important historical and current works on the speed of convergence of empirical quantities is explained in Deheuvels [1990]. The theorems of Donsker and Chung-Kiefer, quoted in Sections B3 and B4, which are concerned with the asymptotic properties of empirical discrepancies for i.i.d. random variables with uniform law on the cube $[0,1]^s$, have been addressed in a number of recent works, which extend the results to independent identically distributed Banach space valued variables, replacing the class of indicator random variables on which the closure is taken with more general classes (see Dudley [1984] and Pollard [1984]). For recent results on the properties of empirical distributions we mention Shorack and Wellner [1986] and Deheuvels and Mason [(1) (2) (3) (4)]. Theorem B.4.2 for the case $s = 1$ is due to Chung [1949]; it has been generalized to other forms [Deheuvels and Mason (1) (2); Shi, 1990]. The case for dimension s was established in Kiefer [1961]; see also Cassels [1951] and Zaremba [1971]. For other extensions

and related problems see Csörgö and Révész [1981] and Gaenssler and Stute [1979].

Section B.5. For more complete presentations of algorithms for stochastic approximation see Benvéniste et al. [1987] and Duflo [1990]. Our account follows Bouleau, Lamberton, and Lapeyre [1986], where the practical importance of stopping conditions is emphasized. For more general assumptions than those of Theorem B.5.1, see Métivier [1983], Benvéniste et al. [1987], and Métivier and Priouret [1987]. The stopping condition B.5.4 is taken from Dvoretzky [1956]. For the speed of algorithms of Robbins–Monro type see also Komlós and Révész [1972], Benvéniste et al. [1987], and Lapeyre [1988].

Algorithms for stochastic approximation have applications in very diverse areas (recursive identification, signal processing, adaptive control), see Benvéniste et al. [1987], as well as for shape recognition and for maximizing a function, as in Kiefer and Wolfowitz [1952], which has been followed by a number of works in the spirit of "simulated annealing algorithms."

There is continuing research aimed at using arithmetic sequences in these algorithms; see Niederreiter [1983, 1988], Niederreiter and McCurley [1979], and Lapeyre, Pagès, and Sab [1990], or in "self-organising" algorithms, see Bouton et al. [1992].

Sections C.1 and C.2. For a survey of low-discrepancy sequences as of 1978, see Niederreiter [1978].

We have given a simplified form of the Koksma–Hlawka inequality [Koksma, 1942; Hlawka, 1961]; the idea for our presentation was suggested to us by G. Pagès. A more elaborate version, which takes into account variation in the sense of Hardy and Krause, is given in Kuipers and Niederreiter [1974], on which we base Section C2.

Section C.3. This section is standard material from ergodic theory. For the proof of Remark C.3.5, see Cornell et al. [1982].

Section C.4. Lapeyre–Pagès sequences are studied in Lapeyre and Pagès [1989]. Faure sequences [H. Faure, 1982] are an improvement upon ideas of Sobol' [1967] and have been further developed; see Niederreiter [1987] and H. Faure [1990]. For more recent developments see Borel et al. [1992].

Sections C.5 and C.6. The circumstances for using low-discrepancy sequences in simulation, for finite dimensions using the rejection method, are outlined in Bouleau [1986].

The speed of deterministic low-discrepancy sequences depends on the dimension as shown and explained in this chapter, but the same applies to random (Borel) sequences, notwithstanding the asymptotic theorem of Chung–Kiefer, as shown by G. Pagès (private communication).

Section C.6 is based on Bouleau [9].

Section D.1. This is standard material. For a characterization of conditional expectations for linear operators from L^p to L^p see Neveu [1972].

Section D.2. Expansion as orthogonal polynomials is a very general method for approximation in L^2, see Nikiforov and Ouvarov [1983] and Bender and Orszag [1978]. For other variations on the algorithm of conjugate gradient see Joly [1988], and for extension to the infinite-dimensional case see Patterson [1974].

Sections D.3 and D.4. This section is based on Bouleau [1986 (2)]. The fact that the methods using polynomials do not necessarily preserve positivity, unlike Haar bases, which are used for pointwise calculations, is shown in Neveu [1972, Chapter II, §3].

Section D.4 is based on Bouleau et al. [1989].

CHAPTER 3

Accessible accounts of the theory of stochastic processes from the point of view of applications can be found in Dacunha-Castelle and Duflo [(1) (2)], Soize [1986], and Bouleau [1988].

Sections A.1 to A.6. These sections largely follow Bouleau [(5) (6)]. For other results, which are sometimes easier to use, and bounds for speed in the shift method, see Ben Alaya [1992].

Section A.7. For more detailed investigations of p-uniformly distributed sequences, see Rauzy [1976] and Bass [1984]. The proof of the proposition concerning Weyl polynomials in A.7.2 is taken from Rauzy [1976].

Section A.8. The theory of Dirichlet forms and their probabilistic interpretation comes from Fukushima [1980] for the finite-dimensional case and Bouleau and Hirsch [1991] for the general case.

Sections B.1 to B.10. Stationary processes are dealt with in classical texts such as Skorohod [1965], Breiman [1968], Shiryayev [1984], Rosenblatt [1985], Bouleau [1988], Preumont [1990]; for discrete-time processes see Azencott et al. [1984]. For accounts that are oriented more toward applications see Lin [1969], Preumont [1990], and Bouc [1991]. For an extension to processes that may not be stationary but admit a spectral representation see Rao [1986].

Among the many works on the simulation of stationary processes we mention Bernard et al. [1984], Borgman [1969], Borgman and Hudspeth [1979], Shinozuka and Jan [1972], Yang and Lin [1983], and Kettani-Wagner [1986].

Sections C.1 to C.3. One of the more recent works on the deep mathematical theory of Markov processes is Dellacherie, Maisonneuve, and Meyer [1992], which contains the standard references on Markov processes. See also Meyer [1967], Getoor [1975], Williams [1979], Fukushima [1980], Dacunha-Castelle and Duflo [1983], Shiryayev [1984], Bliedtner and Hansen [1986], and Bouleau [1988]. The links between Brownian motion and potential theory (harmonic and subharmonic functions) are elucidated in Port and Stone [1978] and Durrett [1984], from which our treatment is partly drawn.

Sections D.1 to D.8. Processes with independent increments are dealt with in classical texts such as Feller [1966]. Here the accent is placed on the links with convolution semi-groups, for which see Berg and Forst [1975]. For the probabilistic interpretation of Lévy measure, see Bretagnolle [1973].

The mathematical theory of stable processes is as general as that of Gaussian processes; see Marcus and Pisier [1983]. The formulas attained in Lepage et al. [1981] allow approximate simulations to stable processes. For the simulation of stable processes see also Boulanger [1990]. These sections are apparently the first account of the use of subordination in the sense of Bochner [1949] in simulation.

Section D.8 follows Bouleau and Chateau [1989] and Chateau [1990].

Sections E.1 to E.4. Our presentation of point processes follows Neveu [1977]; for an application to traffic models see Bouleau [1988]. For applications in geostatistics and the estimation of laws see Matheron [1970].

CHAPTER 4

References on the theory of Markov processes have been given above in the notes for Chapter 3, Section C, and references for the various sections of Chapter 4 are mentioned in the text. For the links between problems of réduites and the mathematics of finance, see Lamberton and Lapeyre [1991]. We give here a sketch of the historical roots of the theory.

The development of the theory of Markov processes is marked by the increasing generality and refinement of results. The main stages include:

- Kolmogorov's discovery [Kolmogorov, 1931] of precise hypotheses under which it is possible to establish explicit links between Markov processes on R^n and elliptic and parabolic partial differential equations. Bachelier [(1) (2)] had expressed these links in an intuitive fashion.

- The interpretation of Newtonian potential theory in terms of trajectories of Brownian motion, begun by Paul Lévy and Kakutani.
- Hunt's construction of a potential theory starting from a general Markov process (using Choquet's definition of capacity to show the measurability of entrance times). In this way, as Meyer, Bauer and Hunt have shown, potential theories are interpreted in terms of Markov processes.

There is a great number of results after this, with the Blumenthal and Getoor framework of standard processes, the Ray, Knight, Meyer-Walsh and Getoor formalization of the very general framework of right processes, and the study of diffusions, that is, of Markov process solutions of stochastic differential equations, notably by Stroock-Varadhan and Ikeda-Watanabe.

For a more detailed history, and references to the works mentioned above, we refer to Bouleau [1987] and the notes of Dellacherie and Meyer [1983, 1987], as well as Dellacherie, Maisonneuve, and Meyer [1992].

The theory of Markov processes has recently been revived by: (1) the fine study allowed by the discovery that stationariness is given by an appropriate choice of entry and exit laws, see Dellacherie, Maisonneuve, and Meyer [1992]; (2) the developments in infinite dimensions, notably for Ornstein–Uhlenbeck processes in Wiener space, using Malliavin calculus, see Bouleau-Hirsch [1991].

Our treatment in Section A.3 of the resolvent family indexed by functions comes from Neveu [1972]; see also Revuz [1975].

For applications, the book by Kree and Soize [1983] applies the techniques of markovianization of stationary processes to the calculus of structures under random actions.

CHAPTER 5

Sections A.1 to A.4. Section A is classical; for more detailed accounts see Revuz and Yor [1991], Protter [1990], Stroock and Varadhan [1979], Ikeda and Watanabe [1981].

Sections B.1 to B.4. In Section B we have given only the simpler results about the discretization of an EDS. This problem has been the subject of much work; see Talay [1990], O. Faure [1990], Bouleau and Talay [1992], Kloeden and Platen [1992]. For an instructive account see Gard [1988].

Sections C.1 to C.4. Section C.1 follows Nualart and Zakai [1990], where deeper results on multiple integrals can be found. The general char-

acter of Lévy's area comes from the fact that the more complicated iterated integrals are expressed continuously in terms of it. The joint law of Lévy's area and the position of a Brownian particle at time 1 was derived by Paul Lévy [1951, 1965]. An integration by parts yields the joint law of $(\omega_1(t), \omega_2(t), \int_0^t \omega_1(s)\, d\omega_2(s), \int_0^t \omega_2(s)\, d\omega_1(s))$. See Gaines and Lyons [1992] for a simulation method for these joint laws.

For an EDS that does not satisfy the commutativity condition, Clark and Cameron [1980] showed that taking values of the Brownian motion at time intervals h only gives an accuracy of $O(\sqrt{h})$. Faster methods are possible using domain integrals (see Pardoux and Talay [1985] and Kloeden and Platen [1992]) which can be simulated approximately or exactly.

These irregularity properties of Brownian functionals provide motivation for attempts to obtain laws for random variables from stochastic calculus; see, for example, Yor [1992].

The property of approximate continuity at regular points of solutions of an EDS in the sense of Stratonovich was discovered by Stroock and Varadhan [1972]. See Ikeda and Watanabe [1981]. For improvements of this result and connected work see Carmona and Nualart [1991], Hargé [1993], Sugita [1992], and Fang [1991]. For a more detailed account than that of Section C.4, see Bouleau and Talay [1992].

Sections D.1 to D.5. We thank G. Caplain and W. Kendall for useful discussions on the definition of simulatable functionals.

Sections E.1 to E.4. For the use in information science of the formal aspects of Ito calculus see Kendall [1991]. For a weakening of the hypotheses of proposition E.1.2 see Meyer [1976]. Our account of the formulas of Hu–Meyer follows Hu and Meyer [(1) (2)].

Formal calculus for stochastic differential equations of Stratonovitch type follows the same rules as for ordinary differential equations. The equations can be solved formally, and the validity of the results can be established under certain assumptions; see Fliess and Norman-Cyrot [1982] and Hu [1990].

The explanation of the Isobe–Sato formula is based on Léandre and Meyer [1989]. Section E.4 draws from the accounts of Azencott [1982] and Ben Arous [1989], which contain conditions for validity of the expansions.

Sections F.1 to F.4. These sections are based on Bouleau [1991] and are the result of discussions with J. P. Conze, E. Lesigne, and M. Ben Alaya. Representations of Brownian motion such as (F.4.3) are of great use for a number of detailed studies of Wiener space. They allow us to see Wiener space as a sequence space; see Ciesielski [1960], Baldi et al. [1992] and Kerkyacharian et al. [1991].

References

References quoted with an asterisk (*) possess an extended bibliography, which can be usefully consulted.

Alfaro M. et al., *Orthogonal polynomials and their applications*, Lecture Notes in Math. 1329, Springer (1988).

Altman N.S., Bit-wise behavior of random numbers generators, *SIAM J. Sci. Stat. Comp.*, 9, No. 5, 941–949 (1988).

Anderson S., Random number generators on super-computers and other advanced architectures, *SIAM Rev.*, 33, no. 2, 221–251 (1990).

Azencott R., Formule de Taylor stochastique et développement asymptotique d'intégrales de Feynmann, in Sem. Prob. XVI supp. Géométrie diff. stoch. Lecture Notes in Math. 921, Springer (1982).

Azencott R. and Dacunha-Castelle D., *Séries d'observations irrégulières*, Masson (1984).

Bachelier L. (1), Theorie de la spéculation, *Ann. Sci. Ecole Normale*, 17, 21–86 (1900).

Bachelier L. (2), Théorie mathématique du jeu, *Ann. Sci. Ecole Normale*, 18, 143–210 (1901).

Baker A., On some diophantine inequalities involving the exponential function, *Can. J. Math.*, 17, 616–626 (1969).

Baldi P., Ben Arous G., and Kerkyacharian G., Large deviations and Strassen theorem in Hölder norm, *Stoch. Proc. and Their Appl.* 42, 171–180 (1992).

Bass J., *Fonctions de corrélation, fonctions pseudo-aléatoires et applications*, Masson (1984).

Bauer H., *Harmonische Raüme und ihre Potentialtheorie*, Lecture Notes in Math. 22, Springer (1966).

Ben Alaya M., Sur les méthodes ergodiques en simulation, thèse ENPC (1992).

Ben Arous C., Flots et séries de Taylor stochastiques, *Prob. Theory Relat. Fields*, 81, 29–77 (1989).

Bender C.M. and Orszag S.A., *Advanced mathematical methods for scientists and engineers*, McGraw-Hill (1978).

Benvéniste A., Métivier M., and Priouret P., *Algorithmes adaptatifs et approximation stochastique, théorie et applications*, Masson (1987).

Berg C. and G. Forst, *Potential theory on locally compact groups*, Springer (1975).

Berger E., An almost sure invariance principle for stationary ergodic sequences of Banach space valued random variables, *Prob. Theory Relat. Fields*, 84, 161–201 (1990).

Bernard P., Méthodes mathématiques d'étude des vibrations aléatoires et analyse sur les espaces gaussiens, Thèse Université Blaise Pascal, Clermont-Ferrand, France (1990).

Bernard P., Bressolette P., Fogli M., and Lemaire M., Un algorithme de simulation stochastique par markovianisation approchée. Application à la mécanique aléatoire, *J. Méc. Th. Appl.*, 3, No. 6, 905–950, (1984).

Billingsley P., *Convergence of probability measures*, Wiley (1968).

Blietner J. and Hansen W., *Potential theory, an analytic and probabilistic approach to balayage*, Springer (1986).

Bochner S., Diffusion equations and stochastic processes, *Proc. Natl. Acad. Sci. USA* 35, 369–370 (1949).

Bonvalot F., Approximation forte de processus empiriques et applications, Thèse Univ. Paris Sud Orsay (1991).

Bony J.M., Courrege Ph., and Priouret P., Semigroupes de Feller sur une variété à bord compacte, *Ann. Inst. Fourier*, 18, 369–521 (1968).

Borel E., *Traité de calcul des probabilités et de ses applications*, Gauthier-Villars (1958).

Borel J.P., Pagès G. and Xiao Y., Suites à discrepance faible et intégration numérique, in Probabilités numériques, Bouleau N. and Talay D. eds., 7-22, INRIA (1992).

Borgman L.E., Ocean waves simulation of engineering design, J. of Waterways and Harbor Division, *Proc. Am. Soc. Civil Eng.*, WW4, 557–583 (1969).

Borgman L.E. and Hudpeth R.T., Efficient FFT simulation of digital times sequences, *J. Eng. Mech. Div.*, 5, No. EM2 (1979).

Borwein J.M. and Borwein P.B., *Pi and the AGM*, Canadian Math. Soc. Series of Monographes and ado-texts, Wiley-Interscience (1987).

Bouaziz L., Mémoire de DEA, Université Paris 6–CERMA-ENPC (1991).

Bouc R. (ed.), Contrôle actif vibro-acoustique et dynamique stochastique Publ. du L.M.A. No. 127, CNRS (1991).

Boulanger F., Modélisation et simulation de variables régionalisées par des fonctions aléatoires stables, Thèse École des Mines de Paris (1990).

Bouleau N. (1), Propriétés d'invariance du domaine du générateur étendu d'un processus de Markov, Sém. Prob. XV, Lecture Notes in Math. 850, Springer (1981).

Bouleau N. (2), *Probabilités de l'ingénieur, variables aléatoires et simulation*, Hermann (1986).

Bouleau N. (3), La jonction entre la théorie du potentiel et les probabilités, Cahiers du Sém. d'Histoire des Mathématiques, Université P. et M. Curie et École Pratique des Hautes Etudes, No 8, 43–66 (1987).

Bouleau N. (4), *Processus stochastiques et applications*, Hermann (1988).

Bouleau N. (5), On effective computation of expectations in large or infinite dimension, *J. Comput. Appl. Math.*, 31, 23–24, (1990).

Bouleau N. (6), On numerical integration by the shift and application to Wiener space, *Acta Appl. Math.*, 25, 201–220, (1991).

Bouleau N. (7), La méthode du balayage de Poincaré comme algorithme numérique: application à la géodésie, *Les Ann. Maghrebines de l'Ing.*, 5, No. 2 (1991).

Bouleau N. (8), Calculs d'espérance par simulation, Matapli (1993).

Bouleau N. (9), A remark on random and equidistributed sequences, *J. Potential Anal.*, 1, 379–384 (1992).

Bouleau N. (10), Suite de points d'un espace de Hilbert distribuée selon la promesure gaussienne canonique, *Note C.R. Acad. Sci. Paris*, 315, sI, 1297–1300 (1992).

Bouleau N. (11), An implementation of irrational translation on the torus, *J. Comput. Appl. Math.* (to appear).

Bouleau N. and Chateau O., Le processus de la subordination, *Note C.R. Acad. Sci. Paris*, 309, sI, 625–628 (1989).

Bouleau N., Chevallier D., Jaillet P., and Lerbet J., Pesage dynamique des véhicules, Contrat LCPC-CERMA (1989).

Bouleau N. and Hirsch F., *Dirichlet forms and analysis on Wiener space*, de Gruyter (1991).

Bouleau N. and Lamberton D., Residual risks and hedging strategies in Markovian markets, *Stoch. Processes Appl.*, 33, 131–150 (1989).

Bouleau N., Lamberton D., and Lapeyre B., Une méthode numérique adaptée au calcul probabiliste des structures, *J. Theor. Appl. Mech.*, 5, No. 5, 781–801 (1986).

Bouleau N., Pagès G., and Xiao Y., Extension des Méthodes de Monte Carlo accélérées, rapport de recherche du contrat DRET/CERMA (opération 88/142) (1990).

*Bouleau N. and Talay D. (eds.), *Probabilités numériques*, INRIA (1992).

Bourbaki N. (1), *Eléments de Mathématiques, Topologie générale*, Hermann (1965).

Bourbaki N. (2), *Eléments de Mathématiques, Intégration*, Hermann (1965).

Bouton C., Cottrell M., Fort J.C., and Pagès G., Self-organization and convergence of the Kohonen algorithm, in Probabilités numériques, Bouleau N. and Talay D. eds., INRIA (1992).

Breiman L., *Probability*, Addison-Wesley (1968).

Brelot M. (1), *Axiomatique des fonctions harmoniques*, Les Presses de l'Université de Montréal (1966).

Brelot M. (2), *Eléments de la théorie classique du potentiel*, Centre de documentation universitaire, Paris (1959).

Bretagnolle J.L., Processus à accroissements indépendants, Lecture Notes in Math. 307, Springer (1973).

Buckley J.A., *Large deviation techniques in decision, simulation and estimation*, Wiley (1990).

Cambanis C. and Soltani A.R., Prediction of stable processes: spectral and moving average representations, *Z. Wahrscheinlichkeitsth. Verw. Geb.*, 66, 593–612 (1984).

Cameron R.H. and Martin W.T., Transformation of Wiener integrals, *Trans. Am. Math. Soc.*, 58, 184–219 (1945).

Caplain G., Propriété de variables effectivement simulables (unpublished communication) (1990).

Carmona R. A. and Nualart D., Trace of random variables on Wiener space and Onsager-Machlup functionals, preprint (1992).

Cassels J.W.S., An extension of the law of iterated logarithm, *Proc. Camb. Philos. Soc.*, 47, 55–64, (1951).

Castelle N., Principe l'invariance et application à la statistique de modèles censurés, Thèse Univ. Paris Sud Orsay (1991).

Chambers J.M., Mallows C.L. and Stuck B., A method of simulating stable random variables, *J. Am. Stat. Assoc.*, 71, No. 354, 340–344, (1976).

Chateau O., Quelques remarques sur les processus à accroissements indépendants stationnaires et la subordination au sens de Bochner, Thèse Univ. Paris 6 (1990).

Choquet G. (1), Répartition des nombres $(3/2)^n$, *C.R. Acad. Sci. Paris*, 290, p. 575ff, p. 719ff, p. 813ff, (1980); 291, p. 69ff, p. 239ff, (1980); 292, p. 5ff, p. 339ff (1981).

Choquet G. (2), Nombres 3/2-constructibles, *Sem. Init. Anal.* No. 9, p. 63ff, (1984/85).

Chung K.L., An estimate concerning the Kolmogorov limit distribution, *Trans. Am. Math. Soc.*, 67, 36–50, (1949).

Ciesielski Z., On the isomorphism of spaces H_α and m, *Bull. Acad. Pol. Sci.*, Ser. Sci. Math. Astron. Phys., VIII, No. 4, 217–222, (1960).

Ciesielski Z. and Taylor S.J., First passage time and sojourn times for Brownian motion in space and exact Hausdorff measure of the sample path, *Trans. Am. Math. Soc.*, 103, 434–450 (1962).

Clark D.S. and Kushner H.J., *Stochastic approximation methods for constrained and unconstrained systems*, Springer (1978).

Clark J.M.C. (1), An efficient approximation scheme for a class of stochastic differential equations. Advances in filtering and optimal stochastic control, Lecture Notes in Control and Inf. Sci. 42, Springer (1982).

Clark J.M.C. (2), The design of robust approximation to the stochastic differential equation of non linear filtering, in *Communication systems and random process theory*, J. Skwirzynskii ed., Sijthoff and Noordhoff (1980).

Clark J.M.C. and R.J. Cameron, The maximum rate of convergence of discrete approximations for stochastic differential equations, in *Stochastic differential system*, Lecture Notes in Control and Inf. Sci. 25, Springer (1980).

Cornfeld P., Fomin S.V. and Sinai Ya.G., *Ergodic theory*, Springer (1982).

Cottrell M., Fort J-C., and Malgouyres G., Large deviations and rare events in the study of stochastic algorithms, *IEEE Trans. Automat. Control*, AC-28, No. 9, 907–920 (1983).

Csörgő M. and Révész P., *Strong approximation in probability and statistics*, Academic Press (1981).

Dacunha-Castelle D. and Duflo M. (1), *Probabilités et statistiques*, 1. *Problèmes à temps fixe*, Masson (1982).

Dacunha-Castelle D. and Duflo M. (2), *Probabilités et statistiques*, 2. *Problèmes à temps mobile*, Masson (1983).

Dagpunar J., *Principles of random variate generation*, Oxford Sci. Pub. (1987).

Deheuvels P., *La Probabilité, le Hasard et la Certitude*, P.U.F. (1990).

Deheuvels P. and Mason D.M. (1), Bahadur–Kiefer type processes, *Ann. Prob.*, 18, 669–691, (1990).

Deheuvels P. and Mason D.M. (2), Nonstandard functional laws of the iterated logarithm, *Ann. Prob.*, 18, 1693–1722 (1990).

Deheuvels P. and Mason D.M. (3), A tail empirical process approach to some non standard laws of iterated logarithm, *J. Theor. Prob.*, 4, 53–85, (1991).

Deheuvels P. and Mason D.M., (4), Functional laws of the iterated logarithm for the increments of empirical and quantile processes, *Ann. Prob.*, 20, 1248–1287, (1992).

Dellacherie C., Nombres au hasard de Borel à Martin Löf, *Gazette des Math.* No. 11, (1978).

Dellacherie C. and Meyer P.A. (1), *Probabilités et Potentiel, chapitres V à VIII, Théorie des martingales*, Hermann (1980).

Dellacherie C. and Meyer P.A. (2), *Probabilités et Potentiel, chapitres IX à XI, Théorie discrète du potentiel*, Hermann (1983).

Dellacherie C. and Meyer P.A. (3), *Probabilités et potentiel, chapitres XII à XVI, Théorie du Potentiel associée à une résolvante Théorie des processus de Markov*, Hermann (1987).

Dellacherie C., Maisonneuve B., and Meyer P.A., Probabilités et Potentiel, *chapitres XVII à XXIV, Processus de Markov (fin), Compléments de calcul stochastique*, Hermann (1992)

*Devroye L., *Non-uniform random variate generation*, Springer (1986).

Doss H. Liens entre équations différentielles stochastiques et ordinaires, *Ann. Inst. H. Poincaré*, XIII, 2, 99–125 (1977).

Doob J.L., *Stochastic processes*, Wiley (1953).

*Dudley R.M., A course on empirical processes, in Ecole d'Eté de Probabilité de Saint Flour XII-1982, Lecture Notes in Math. 1097, Springer (1984).

Duflo M., *Méthodes récursives aléatoires*, Masson (1990).

Dumont J.M. and Thomas A. Une modification des nombres g-normaux. *Ann. Fac. Sci. Toulouse*, III, 367–373 (1986).

Dupain Y. and Sós V.T., On the one-sided boundedness of discrepancy functions of the sequence $\{n\alpha\}$, *Acta Arith.* 37, 363–374 (1980).

Durrett R., *Brownian motion and martingales in analysis*, Wadsworth (1984).

Dvoretzky A., On stochastic approximation. *Proc. 3rd Berkeley Symp. on Math. Stat and Prob.*, vol. 1, 35–56, (1956).

El Karoui N., Les aspects probabilistes du controle stochastique, in Lecture Notes in Math. 876, 72–238, Springer (1981).

Ermakov S.M., Nekrutkin V.V., and Sipin A.S., *Random processes for classical equations of mathematical physics*, Kluwer (1989).

Fang Sh., Sur la continuité approximative forte des fonctionnelles de Ito, *Stochastics Stoch. Rep.*, 36, 193–204, (1991).

Faure H. (1), Discrépance de suites associées à un système de numération (en dimension s), *Acta Arith.*, 41, 337–351, (1982).

Faure H. (2), Using permutation to reduce discrepancy, *J. Comput. Appl. Math.*, 31, 97–103, (1990).

Faure O. (1), Numerical pathwise approximate of stochastic differential equations. CERMA (1990).

*Faure O. (2), Simulation du mouvement brownien et des diffusions, Thèse ENPC, Paris (1992).

Faure O. and Gaines J., Simulation trajectorielle des diffusions, in Probabilités numériques, Bouleau N. and Talay D. eds., INRIA (1992).

Félix L., *Message d'un mathématicien:* Henri Lebesgue, A. Blanchard (1974).

Feller W., *Introduction to probability theory and its applications*, Vol. II, Wiley (1966).

Feyel D., Propriétés de permanence d'un générateur infinitésimal, in Sém. Th. Potentiel No. 4, 38–50, Lecture Notes in Math. 713, Springer (1979).

Feyel D. and A. de la Pradelle, Étude de l'équation $\frac{1}{2}\Delta u - u\mu = 0$ où μ est une mesure positive, *Ann. Inst. Fourier*, 38, 199–218, (1988).

Finkelstein H., The law of the iterated logarithm for empirical distributions, *Ann. Math. Stat.*, 42, 607–615, (1971).

Fishman G.S. and Moore L.R., An exhaustive analysis of multiplicative congruential generators with modulus $2^{31} - 1$, *SIAM J. Sci. Statist. Comput.*, 7, 24–45, (1986).

Flajolet P., Kirchenhofer P. and Tichy R.F., Deviation from normality in random strings, *Prob. Th. Rel. Fields*, 80, 139–150, (1988).

Fliess M. and Norman-Cyrot D., Algèbres de Lie nilpotentes, formules de Campbell-Baker-Hausdorff et intégrales itérées de Chern, Sém. Prob. XVI, Lecture Notes in Math. 920, 257–267, Springer (1982).

Friedman A., *Stochastic differential equations and application*, Vol. 1, Academic Press (1976).

Fukushima M., *Dirichlet forms and Markov processes*, North-Holland Kodansha (1980).

Fushimi M., Designing a uniform random number generator whose subsequences are k-distributed, *SIAM J. Comput.*, 17, No. 1, 89–99 (1988).

Gaenssler P. and Stute W., Empirical processes: a survey of results for independent and identically distributed random variables, *¡Ann. Prob.*, 7, No. 2, 193–243 (1979).

Gaines J. and Lyons T.J., Random generation of stochastic area integrals, *Proceedings of the 1st Workshop on Stochastic Numerics*, Berlin (1992).

Gard T.C., *Introduction to stochastic differential equations*, Dekker (1988).

Getoor R.K., *Markov processes: Ray processes and right processes*, Lecture Notes in Math. 440, Springer (1975).

Gordin M.I., The central limit theorem for stationary sequences, *Sov. Math. Dokl.*, 10, No. 5, 1174–1175 (1969).

Guégan D., Modèles bilinéaires et polynomiaux de séries chronologiques: étude probabiliste et analyse statistique, Thèse Univ. Grenoble I (1988).

Hahn P. and Jeruchim M., Developments in the theory and application of importance sampling, *IEEE Trans. Commun.*, COM-35, No. 7, 706–714, (1987).

Hall P. and Heyde C.C., *Martingale limit theory and applications*, Academic Press (1980).

Halász G., Remarks on the remainder in Birkhoff's ergodic theorem, *Acta Math. Acad. Hungar.*, 28,389–395 (1976).

Halmos P.R., Applied mathematics is bad mathematics, in *Mathematics tomorrow*, Steen L. A. ed., Springer (1981).

Halton J.H., On the efficiency of certain quasi-random sequences of points in evaluating multidimensional integrals, *Numer. Math.*, 2, 84–90 (1960).

Hammersley J.M., Monte Carlo methods for solving multivariate problems, *Ann. N.Y. Acad. Sci.*, 86, 844–874 (1960).

Hargé G., Régularité de certaines fonctionnelles sur l'espace de Wiener, Thèse Univ. Evry val d'Essonne (1993).

Harrison J. and Pliska S., Martingales and stochastic integrals in the theory of continuous trading, *Stoch. Processes Appl.*, 11, 215–260 (1981).

Hecke E., Uber analytische Functionen und die Verteilung von Zahlen mod Eins, *Abh. Math. Sem. Hamburg*, 1, 54–76 (1922)

Heyde C.C. and D.J. Scott, Invariance principles for the law of iterated logarithm for martingales and processes with stationary increments, *Ann. Prob.*, 1, No. 3, 428–436 (1973).

Hille E. and Phillips R.S., *Functional analysis and semigroups*, American Mathematics Society (1957).

Hirsch F. (1), Intégrales de résolvantes et calcul symbolique, *Ann. Inst. Fourier*, XXII, 4, 239–264 (1972).

Hirsch F. (2), Opérateur carré du champ [d'après Roth], in Sém. Bourbaki (1976/77).

Hlawka E., Funktionen von beschränkter Variation in der Theorie Gleichverteilung, *Ann. Math. Pure Appl.*, 54, 325–333 (1961).

Hu Y.Z., Calculs formels sur les EDS de Stratonovitch, Sém. Prob. XXIV, Lecture Notes in Math. 1426, 453–460, Springer (1998).

Hu Y.Z. and Meyer P.A. (1), Chaos de Wiener et intégrales de Feynman, Sém. Prob. XXII, 51–71, Lecture Notes in Math 1321, Springer (1988).

Hu Y.Z. and Meyer P.A. (2), Sur les intégrales multiples de Stratonovitch, Sém. Prob. XXII, 72–81, Lecture Notes in Math 1321, Springer (1988).

Hua Loo Keng and Wang Yuan, *Applications of number theory to numerical analysis*, Springer (1981).

*Ikeda N. and S. Watanabe, *Stochastic differential equations and diffusion processes*, North Holland (1981).

Isobe E. and Sato Sh., Wiener-Hermite expansion of a process generated by an Ito stochastic differential equation, *J. Appl. Prob.*, 20, 754–765 (1983).

Ivanyi A., Construction of infinite de Bruijn arrays, *Discrete Appl. Math.*, 22, 289–293 (1988/89).

Jacod J., *Calcul stochastique et problèmes de martingales*, Lecture Notes in Math. 714, Springer (1979).

Joly P., Présentation de synthèse des algorithmes de gradient conjugué, *Math. Model. Num. Anal. 20,*, No. 4, 639–665 (1988).

Kac M., *Enigmas of chance*, University of California Press (1987).

Kendall W., Symbolic Ito calculus: an overview, in Probabilités numériques, Bouleau N. and Talay D. eds., INRIA (1991).

Kerkyacharian G. and Roynette B., Une démonstration simple des théorèmes de Kolmogorov, Donsker et Nisio, *C.R. Acad. Sci. Paris*, 312, 877–882 (1991).

Kettani-Wagner S., Étude comparative de techniques de simulation de fonctions aléatoires gaussiennes appliquée à l'analyse de la sécurité des structures, Thèse Univ. Clermont II (1986).

Kiefer J., On large deviation of the empiric D.F. of vector chance variables and a law of iterated logarithm, *Pacific J. Math.*, 11, 649–660 (1961).

Kiefer J. and Wolfowitz J. Stochastic estimation of the modulus of a regression function. *Ann. Math. Stat.*, 23, 462–466 (1952).

Kloeden P.E. and Platen E., *Numerical solution of stochastic differential equations*, Springer (1992).

Knight F.B., *Essential of Brownian motion*, Math. Survey, 18, AMS (1981).

Knuth D., *The art of computer programming*, Vol. 2, semi-numerical algorithms, Addison-Wesley (1981).

Knuth D. and Yao A., The complexity of non uniform random generation, in Algorithms and complexity, J.F. Taub ed., 357–429, Academic Press (1976).

Koksma I.F., Een algemeene stelling uit de theorie der gelikmatige Verdeeling modulo 1, *Math. B (Zutphen)*, 11, 7–11 (1942/43).

Kolmogorov A.N., *Problèmes de théorie de l'information*, Acad. Sci. URSS, Moscou (1969).

Kolmogorov A.N., Uber die analytischen Methoden in der Wahrscheinlichkeitsrechnung, *Math. Ann.*, 104, 415–458, (1931).

Komlós J., Major P., and Tusnady G., An approximation of partial sums of independent R.V.'s and the sample D.F.'s, *Z. Wahrscheinlichkeitsth. Verw. Geb.*, 32, 111–121 (1975).

Komlós J. and Révész P., On the rate of convergence of the Robbins–Monro method, *Z. Wahrscheinlichkeitsth. Verw. Geb.*, 25, 39–47 (1972).

Kôno N., 4-Dimensional Brownian motion is recurrent with positive capacity, *Proc. Jpn. Acad.*, 60, 57–59 (1984)

Kowada M., Convergence rate in the ergodic theorem for an analytic flow on the torus, Lecture Notes in Math. 330, 251–254, Springer (1973).

Krengel U. (1), On the speed of convergence in the ergodic theorem, *Monatsh. Math.*, 86, 3–6 (1978).

Krengel U. (2), *Ergodic theorems*, de Gruyter (1985).

Kree P. and Soize Ch., *Mécanique aléatoire*, Dunod (1983).

*Kuipers L. and H. Niederreiter, *Uniform distribution of sequences*, Wiley (1974).

Kunita H., Stochastic differential equations and stochastic flows of diffeomorphisms. in Ecole d'été de St Flour XII, 143–303, Lecture Notes in Math. 1087, Springer (1982)

Kushner H.J., *Probability methods for approximations in stochastic control and for elliptic equations*, Academic Press (1977).

Lamberton D. and Lapeyre B., *Introduction au calcul stochastique appliqué à la finance*, SMAI, Ellipses (1991).

Lamberton D. and Pagès G., Sur l'approximation des réduites, *Ann IHP, Sér. Prob. Stat.*, 26, No. 2, 331–355 (1991).

Lapeyre B., Etude de quelques questions relatives aux processus de diffusion issues de la mécanique aléatoire, Thèse Univ. Paris VI (1988).

Lapeyre B. and Pagès G., Familles de suites à discrépance faible obtenues par itération de transformations de [0,1], *C.R. Acad. Sci.*, 308, sI, 507–509 (1989).

Lapeyre B., Pagès G., and Sab K., Sequences with low discrepancy. Generalisation and application to Robbins-Monro algorithm, *Statistics*, 21, No. 2, 251–272 (1990).

Laplace P. S., *Essai philosophique sur les probabilités*, Gauthier-Villars, (1921).

Léandre R. and Meyer P.A., Sur le développement d'une diffusion en chaos de Wiener, Sém. Prob. XXII, 161–164, Lecture Notes in Math. 1372, Springer (1989).

Lemarié P.G. and Meyer Y., Ondelettes et bases hilbertiennes, *Rev. Mat. Ibero Am.*, 2, No. 1-2 (1986).

Lepage R., Woodroofe M., and Zinn J., Convergence to stable distribution via order statistics, *Ann. Prob.*, 9, 624–632 (1981).

Lépingle D. and Ribémont B., Un schéma multipas d'approximation de l'équation de Langevin, *Stoch. Proc. and Their Appl.*, 37, 61–69 (1991).

Lévy P., Wiener's random function and other Laplacian random functions, 2d Berkeley Symp. on Math. Stat. and Prob., Neyman J. ed. (1951).

Lévy P., *Processus stochastiques et mouvement brownien*, Gauthier-Villars (1965).

Lévy P., *Quelques aspects de la pensée d'un mathématicien*, A. Blanchard (1970).

Lhote A., *Les invariants plastiques*, Hermann (1967).

Lin Y.K., *Probabilistic theory of structural dynamics*, Krieger (1967).

Marcus B. and Pisier G., Characterization of almost surely continuous p-stable random Fourier series and strongly stationary processes, *Acta Math.*, 152, 245–301 (1983).

Marsaglia G., A current view of random numbers generators, in *Computer science and statistics*: The interface, 3–10, Elsevier Science Publishers (1985).

Marsaglia G. and Zaman A., Toward a universal random number generator, *Stat. Prob. Lett.*, 8, 35–39 (1990).

Martin-Löf P., The definition of random sequences, *Inf. Control*, 9, No. 6, 602–619 (1966).

Maruyama G., Continuous Markov processes and stochastic equations, *Rend. Circ. Mat. Palermo*, 4, 48–90 (1955).

Matheron G., La théorie des variables régionalisées et ses applications, Les Cahiers du Centre de Morphologie Math., Ecole des Mines, Fontainebleau (1970).

McShane E.J., *Stochastic calculus and stochastic models*, Academic Press (1974).

Meijer H.G., The discrepancy of a g-adic sequence, *Indag. Math.*, 30, 54–66 (1968).

Métivier M., *Semimartingales*, de Gruyter (1983).

Métivier M. and Priouret P., Théorèmes de convergence presque-sûre pour une classe d'algorithmes stochastique à pas décroissants, *Prob. Theory Relat. Fields*, 74, 403–428 (1987).

Metropolis, N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., and Teller E., Equations of state calculations by fast computing machines, *J. Chem. Phys.*, 21, 1087–1092 (1953).

Meyer P.A. (1), *Probabilités et Potentiel*, Hermann (1966).

Meyer P.A. (2), *Processus de Markov*, Lecture Notes in Math. 26, Springer (1967).

Meyer P.A. (3), Intégrales stochastiques I-IV, Sém. de Probabilités de Strasbourg I, Lecture Notes in Math. 39, 72–162, Springer (1967).

Meyer P.A. (4), Démonstration probabiliste de certaines inégalités de Littlewood-Paley, Exposé II, Sém. Prob. X, Lecture Notes in Math. 5511, Springer (1976).

Meyer P.A. (5), Un cours sur les intégrales stochastiques, Sém. Prob. X, Lecture Notes in Math. 511, Springer (1976)

Milstein G.N., Approximate integration of stochastic differential equations, *Theory Probab. Its Appl.* (*USSR*), 19, 557–562 (1974).

Milstein G.N., A method of second order accuracy of stochastic differential equations, *Theory Probab. Its Appl.* (*USSR*), 23, 396–401 (1976).

Mogul'skii A.A., On the law of iterated logarithm in Chung's form for functional spaces, *Theory Probab. Its Appl.* (*USSR*), 24, 405–413 (1979).

Mokobodzki G. (1), Quelques propriétés remarquables des opérateurs presque positifs, Sém. Prob. IV, Lecture Notes in Math. 124, Springer (1970).

Mokobodzki G. (2), Cônes de potentiel et noyaux subordonnés, in Potential theorie, CIME Stresa 1969, Ed. Cremonese, 207–248 (1970).

Mokobodzki G. (3), Sur l'algèbre contenue dans le domaine étendu d'un générateur infinitésimal, Sém. Th. Potentiel No. 3, Lecture Notes in Math. 681, Springer (1978).

Mokobodzki G. and Sibony D., Cones de fonctions et théorie du potentiel, in Sém. de Théorie du Potentiel Brelot Choquet Deny, I, 8.01–8.35 (1968); II, 9.01–9.29 (1968).

Morgan, B.J.T., *Elements of simulation*, Chapman and Hall (1984).

Morillon B., Both J.P., and Nimal J.C., Simulation du transport de particules non chargées, 117ème congrès nat. des Soc. savantes (1992).

Neveu J., Martingales à temps discret, Masson (1972).

Neveu, J., Potentiel markovien récurrent des chaines de Harris, *Am. Inst. Fourier* 22, 85–130 (1972).

Neveu J., Processus ponctuels, Ecole d'été de St Flour, Lecture Notes in Math. 598, Springer (1977).

Newton N.J., An asymptotic efficient difference formula for solving Ito stochastic differential equations, *Stochastics*, 19, 178–206, (1986).

*Niederreiter H. (1), Quasi-Monte Carlo methods and pseudo-random numbers, *Bull. Am. Math. Soc.*, 84, 957–1041 (1978).

Niederreiter H. (2), A quasi Monte Carlo method for the approximate computation of the the extreme values of a function, in Studies in pure mathematics (to the memory of Paul Turán) 523–529, Birkhäuser (1983).

Niederreiter H. (3), Point sets and sequences with small discrepancy, *Monatsh. Math.*, 104, 273–337 (1987).

Niederreiter H. (4), Low-discrepancy and low-dispersion sequences, *J. Number Theory*, 30, No. 1, 51–70 (1988).

Niederreiter H. and McCurley K., Optimization of functions by quasi-random search methods, *Computing*, 22, 119–123 (1979).

Nikiforov A. and Ouvarov V., *Fonctions spéciales de la Physique mathématique*, Ed. de Moscou (1983).

Nualart D. and M. Zakai, Multiple Wiener–Ito integrals possessing a continuous extension *Prob. Theory Relat. Fields*, 85, 134–145 (1990).

Ornstein D.S. and Weiss B., Statistical properties of chaotic systems, *Bull. AMS*, 24, No. 1, 11–116 (1991).

Pagès G. and Xiao Y., Sequences with low discrepancy and pseudo-random numbers: theoretical remarks and numerical tests, (Unpublished report) (1991).

Patterson W.M., Iterative methods for the solution of a linear operator equation in Hilbert space, Lecture Notes in Math. 394, Springer (1974).

Pardoux E. and Talay D., Approximation and simulation of solutions of SDE, *Acta Appl. Math.*, 3, 23–47 (1985).

Philoche C., Mémoire de DEA, Université Paris 6—CERMA (1991).

Pitman J. and Yor M., A decomposition of Bessel bridges, *Z. Wahrscheinlichkeitsth. Verw. Geb.*, 59, 425–457 (1982).

Platen E., An approximation method for a class of Ito processes with jump component, *Liet. Mat. Rink.*, 22, 124–136 (1982).

Platen E., Derivative free numerical methods for stochastic differential equations, Lecture Notes in Control and Inf. Sci., 96, 187-193, Springer (1987).

Poincaré H. (1), *Calcul des Probabilités*, Cours Fac. Sci. Paris 1912, Gabay (1987).

Poincaré H. (2), *Théorie du potentiel newtonien*, cours à la Sorbonne 1894–1895, Carré et Naud (1899).

Pollard D., *Convergence of stochastic processes*, Springer (1984).

Port S.C. and Stone C.J., *Brownian motion and classical potential theory*, Academic Press (1978).

Preumont A., *Vibrations aléatoires et analyse spectrale*, Presses Polytechniques et Univ. Romandes (1990).

Prigogine I., *Physique temps et devenir*, Masson (1980).

*Protter Ph., *Stochastic integration and differential equations*, a new approach, Springer (1990).

Rao M.M. (ed.), *Real and stochastic analysis*, Wiley (1986).

Rauzy G. (1), *Propriétés statistiques des suites arithmétiques*, P.U.F. (1976).

Rauzy G. (2), Discrépance d'une suite complètement équirépartie, *Ann. Fac. Sci. Toulouse Math.*, 3, 105–112 (1981) .

Revuz D., *Markov chains*, North Holland (1975).

Revuz D. and Yor M., *Continuous martingales and Brownian motion*, Springer (1991).

Ripley B.D. (1), *Stochastic simulation*, Wiley (1987).

Ripley B.D. (2), Thoughts on pseudo-random numbers generators, *J. Comput. Appl. Math.*, 31, 153–163 (1990).

Ripley B.D. and Kirkland M.D., Iterative simulation methods, *J Comput. Appl. Math.*, 31, 165–172 (1990).

Robbins M. and Monro S., A stochastic approximation method, *Ann. Math. Stat.*, 22, 400–401, (1951).

Rogers L.C.G. and Williams D., *Diffusions, Markov processes and martingales*, Vol. 2: Ito calculus, Wiley (1987).

Rosenblatt M., *Stationary sequences and random fields*, Birkhaüser (1985).

Rosinski J., On stochastic integral representation of stable processes with sample paths in Banach spaces, *J. Multivariate Anal.*, 20, 277–302 (1986).

Rubinstein R. (1), *Simulation and the Monte Carlo method*, Wiley (1981).

Rubinstein R. (2), *Monte Carlo optimization, simulation and sensitivity of Queuing networks*, Wiley (1986) .

Rudin W., *Real and complex analysis*, McGraw-Hill (1966).

Rümelin W., Numerical treatment of stochastic differential equations, *SIAM J. Num. Anal.*, 19, 604–613 (1982).

Sarkar P.K. and Prasad M.A., A comparative study of pseudo and quasi random sequences for solution of integral equations, *J. Comput. Phys.*, 68, 66–88 (1987).

Schmidt W.M., Simultaneous approximation to algebraic numbers by rationals, *Acta Math.*, 125, 189–201 (1970).

Schwartz L., *Topologie générale et analyse fonctionnelle*, Hermann (1970).

Scott D.J., Central limit theorems for martingales and processes with stationary increments using a Skorokod representation approach, *Adv. Appl. Prob.*, 5, 119–137 (1973).

Shi Z., Loi du logarithme itéré pour le processus empirique associé aux U-statistiques, *C.R. Acad. Sci. Paris*, 310, sI, 751–756 (1990).

Shinozuka M. and Jan C.M., Digital simulation of random processes and its applications, *J. Sound Vibrations*, 52, 111–128 (1972).

Shiryayev A.N., *Probability*, Springer (1984).

*Shorack G. and J. Wellner, *Empirical processes with application to statistics*, Wiley (1980).

Skorohod A.V., *Studies in the theory of random processes*, Addison-Wesley (1965).

Smyrnelis E.P., Axiomatique des fonctions biharmoniques, *Ann. Inst. Fourier*, I, 25, 35–97 (1975); II, 26, 1–47 (1976).

Sobol' J.M. (1), The distribution of points in a cube and approximate evaluation of integrals, *Zh. Vichisl. Mat. Mat. Fiz.*, 7, 784–802 (1967).

Sobol' J.M. (2), *Infinite dimensional uniformly distributed sequences in numerical mathematics* (Russian), Inst. Prikl. Mat. Akad. Nauk. SSR, Moscow (1974).

Sobol' J.M. (3), On the convergence of infinite dimensional cubature and simulation of Markov chains, *Voprosy Vyčisl. Prikl. Mat. (Taškent)* 32, 161–167 (1975).

Soize C., Processus stochastiques et méthodes de résolution des problèmes aléatoires, Cours de l'Ecole Centrale des Arts et Manufactures, (1986).

Stroock D.W. and Varadhan S.R.S., On the support of diffusion processes with application to the strong maximal principle, Proc. 6th Berkeley Symp. Math. Stat. Prob. III, 333–359 (1972).

Stroock D.W. and Varadhan S.R.S., *Multidimensional diffusion processes*, Springer (1979).

Sugita H. Hu-Meyer's multiple Stratonovitch integral and essential continuity of multiple Wiener integral, *Bull. Sci. Math. 2ème Sér.*, 113, 463–474 (1989).

Sugita H., Various topologies in the Wiener space and Lévy's stochastic area, *Prob. Theory Relat. Fields*, 91, 283–296 (1992).

Sussmann H.J. On the gap between deterministic and stochastic differential equations, *Ann. Prob.*, 6, 19–44 (1978).

Talay D. (1), Résolution trajectorielle et analyse numérique des équations différentielles stochastiques, *Stochastics*, 9, 275–306 (1983).

Talay D. (2), Discrétisation d'une équation différentielle stochastique et calcul approché d'espérance de fonctionnelles de la solution, M^2AN, 20, 141–179 (1986).

Talay D. (3), Simulation and numerical analysis of stochastic differential systems: a review, INRIA Rapport 1313 (1990).

Tusnady G., A remark on the approximation of the sample DF in the multidimensional case, *Periodica Math. Hung.*, 8, No. 1, 53–55 (1977).

Ventsel' A.D., Gladyshev S.A., and Milstein G.N., Piecewise constant approximations for Monte Carlo calculation of Wiener integrals, *Theory Probab. Its Appl. (USSR)*, XXIX, 4, 744–752 (1982).

Veretennikov A.Y. and Krylov N.V., On explicit formulas for solutions of stochastic differential equations, *Mat. Sbornik*, 100 (1976); English transl.: *Math. USSR Sbornik*, 29, 239–256 (1976).

Von Neumann J., Various techniques used in connection with random digits, *U.S. Natl. Bur. Stand. Appl. Math. Ser.*, No. 12, 36–38 (1951).

Wagon W., Is π normal?, *Math. Intelligencer*, 7, 3, 65–67 (1985).

Warnock T., Low discrepancy point sets in transport codes, in *Application of number theory to numerical analysis*, Zaremba S.K. ed., Academic Press (1972).

Watanabe S., *Stochastic differential equations and Malliavin calculus*, Tata Institute, (1984).

Wiener N., *Cybernetics*, Hermann (1958).

Wiener N. and Siegel A., *Phys. Rev.*, 91, 1551ff (1953); *Il Cimento Nuovo Suppl.*, 2, No. 4, 982ff (1955).

Williams D., *Diffusions, Markov processes and martingales*, Wiley (1979).

Xiao Y., Contribution aux méthodes numériques pour la simulation accélérée, Thèse Ecole Nat. Ponts et Chaussées (1990).

Yamada T., Sur l'approximation des solutions d'équations différentielles stochastiques, *Z. Wahrscheinlichkeitsth. Verw. Geb.*, 36, 153–164 (1974).

Yang J.N. and Lin M.J., Building critical model control non stationary earthquake, *J. Eng. Mech.*, 109, 6, 1375–1389 (1983).

Young L., *Mathematicians and their times*, North-Holland (1981).

Yarmolik V.N. and Demidenko S.N., *Generation and application of pseudorandom sequences for random testing*, Wiley (1988).

Yor M., *Some aspects of Brownian motion*, Part I, Birkhaüser (1992).

Zaremba S.K., Sur la discrépance des suites aléatoires, *Z. Wahrscheinlichkeitsth. Verw. Geb.*, 20, 236–248 (1971).

Zielinski R., An aperiodic pseudorandom number generator, *J. Comput. Appl. Math.*, 31, 205–210 (1990).

Index

$\bar{A}$, 3
$\mathring{A}$, 3
$A \backslash B$, 2
$A \triangle B$, 2
Accelerated simulation, 100
Additive functional, 237, 241
Algebra, 5, 6
Algebraic manifold, 99
Algebraic number, 2
Algorithm, 100
Algorithms for simulation, 46
Analytic, 292
Approximate continuity property, 287
Approximate discrepancy, 93
Approximate sequence, 80
Approximate simulation, 211
Ascoli's theorem, 5, 74

Baker's transformation, 128
Balayage, 241
Balayage algorithm, 241, 250
Balayée, 243
Banach space, 3, 230, 329
Band-pass filter, 151
Bernoulli shift, 124
Bernoulli shift method, 113
Bernstein function, 162, 202, 204, 205, 213, 219
Bernstein monoid, 219
Bessel function, 198
Birkhoff's theorem, 18, 121, 319
Bochner's theorem, 10, 132, 149
Borel function, 6
Borel measure, 9
Borel sequence, 46, 56, 82, 86, 115, 131
Borel σ-field, 5
Boreloid μ-distributed sequence, 87
Boundary of A, 3
Branch, 219
Brownian bridge, 28, 51, 327
Brownian functional, 283, 284
Brownian motion, 26, 124, 142, 145, 159, 162, 174, 187, 189, 203, 216, 270, 278, 284, 302, 306, 320
Burkholder inequality, 19, 62, 269

Cameron–Martin function, 290
Cameron–Martin–Girsanov theorem, 30, 285
Cameron–Martin space, 32, 286
Capacity, 133, 137
Carathéodory extension theorem, 7, 143, 225
Carré du champ operator, 262
Cauchy process, 203
Cauchy semi-group, 203
Central limit theorem, 13, 50
Chain rule, 25, 306, 309
Champernowne sequence, 84, 124

Chaos expansion, 313
Chebyshev polynomial, 105
Chernov's theorem, 55
Closure of A, 3
c-null set, 138
Compact, 4
Complete, 3
Complete market assumption, 265
Completely dissymmetric process, 210
Complex Gaussian random variable, 147
Complex standard Brownian motion, 159
Conditional expectation given $\mathscr{B}$, 14
Conditional expectation of X given Y, 95
Conditional probability of A given $\mathscr{B}$, 14
Conformal local martingale, 160
Congruence method, 48
Conjugate gradient algorithm, 100
Contingent claim, 265
Convergence-determining set, 35, 141
Convergence in distribution or in law, 13, 38
Convolution, 41, 92, 218
Convolution semi-group, 200, 202, 203, 213
Covariance, 12, 51
Covariance matrix, 13
Cramér transform, 55

∂A, 3
$D_\infty^*(\xi, N)$, 44
$D_\infty(\xi, N)$, 44
$D_p^*(\xi, N)$, 44
$D_p(\xi, N)$, 44
de Bruijn sequence, 124
Density function, 12
Difference of A and B, 2
Differential equation, 153
Dirichlet form, 129, 133
Dirichlet problem, 237,
Dirichlet structure, 133
Discrepancy, 44, 67, 76
Discretization, 167, 172
Discretization of SDE, 274
Distribution, 10, 31
Distribution function, 11, 38, 40
Dirac measure, 8, 217
Domination principle, 246
Doob, 34
Doob–Donsker theorem, 51
Doob–Meyer decomposition, 21
Doob's h-transform, 196
Doob's inequality, 19, 21, 269
Doob's martingale convergence theorem, 19, 21, 107
Dvoretzky, 64
Dyadic presentation, 302
Dynamical weighting, 110

Effectively in L^1, 117
Empirical distribution, 49, 50
Equivalent, 8
Ess sup, 253
Ergodic, 18, 71, 179
Ergodic endomorphism, 125
Ergodic theorem, 18, 121, 319
Ergodic theory, 18
Euclidean space, 4
Euler scheme, 274, 278
Evanescent, 16
Excessive function, 245, 250, 256
Exit time, 193
Expectation, 12
Exponent function, 165
Extended generator, 258, 262

Fatou's lemma, 136
Faure sequence, 77
Feller process, 230, 262
Feller semi-group, 231, 239, 255, 260

Feynman integral, 241
Feynman–Kac formula, 237, 240
Filling operator, 252
Filling scheme, 252
Filtration, 18, 20
Finite measure, 6
Finite variation, 21, 67
Fourier series, 72
Fourier transform, 149
Fubini's theorem (stochastic), 152, 280

Gál and Koksma, 125
Gain, 150
Γ-process, 203, 218
Gaussian integer, 2, 75
Gaussian process, 27
Gaussian random variable, 12
General topology, 3
Generator (infinitesimal), 239, 311
Generator (random number), 46, 299
Generic point, 142
Geometric law, 217
Glivenko–Cantelli theorem, 48
Gordin class, 127, 130, 319
Gordin decomposition, 127, 130
Gronwall lemma, 269, 272, 274, 281

Haar function, 327
Hahn–Banach theorem, 3, 232
Halmos Paul, 262
Halton sequence, 75
Hammersley sequence, 76
Harmonic function, 242, 243
Harmonic kernel, 244
Hartman–Wintner law of the iterated logarithm, 56
Hausdorff space, 4
Heat semi-group, 203
Hedging strategy, 265
Hermite polynomial, 105
Hermitian form, 3
Heyde and Scott, 125
Hilbert space, 3
Hille–Yosida theorem, 239
Hitting time, 22, 122, 245
Hölder condition, 279
Hölder property, 275
Holomorhic function, 202, 300
Homographic branch, 220
Hu–Meyer formula, 306
Hunt's theorem, 236

Indicator function, 3
Indistinguishable, 16
Infinite dimension, 113
Infinitely divisible law, 216
Initial measure, 16
Integration by parts formula, 25
Interior of A, 3
Invariance principle, 53
Inverse function, 49
Irrational, 71, 132
Irrational number, 45
Irrational translations on the torus, 71, 133
Isobe–Sato formula, 310
Iterated stochastic integral, 314
Iterative simulation, 183, 284
Ito–Stratonovich transformation, 306
Ito's formula, 25, 264, 273, 287
Ito's theorem, 270, 311

Jump process, 182, 217
Jump PSII, 204

Kac Marc, 111, 241
Kiefer–Wolfowitz procedure, 58
Knight's theorem, 31, 288
Koksma–Hlawka inequality, 67, 83
Kolmogorov's extension theorem, 15
Kolmogorov–Smirnov theorem, 51

Laguerre polynomial, 105

λ-excessive function, 235
λ-supermedian function, 236
Langevin equation, 279
Lapeyre–Pagès sequence, 76
Laplace Pierre Simon, 54
Laplace transform, 202, 207
Law, 11
Law of the iterated logarithm, 56, 125, 319
Least-squares method, 247
Lebesgue measure, 8
Lebesgue type process, 263
Legendre polynomial, 104
Lévy–Hincin formula, 165, 201, 209
Lévy's area, 287
Lévy measure, 202, 204, 205, 209, 217
Lévy Paul, 187, 293
Lhote André, 300
Lifetime, 17, 233
Linearly independent, 99
Lipschitz hypotheses, 270, 275, 321
Local martingale, 23, 160, 264
Locally compact, 4
Low discrepancy, 75
Low-discrepancy sequence, 67, 109, 228
Lower semi-continuous, 36
lsc, 36, 299
Lusin property, 90

Markov chain, 16, 116, 118, 122, 244
Markov kernel, 97, 182
Markov process, 181, 214, 220, 255, 265, 312
Markov property, 17, 33, 233, 238
Markovian potential theory, 230
Markovian problem, 230
Martingale, 18, 20, 116, 235, 258, 263, 265, 269, 288
$\max(x, y)$, 1
Measurable semi-group, 231
Measurable space, 5
Measure-preserving transformation, 18, 71
Mechanical system, 155
Mesh, 23
Metrizable, 3
$\min(x, y)$, 1
Mixing transformation, 18
Model, 111, 221
Modification, 15
Mokobodzki, G., 253
Monotone class, 6
Monotone class theorem, 6, 233
Monte Carlo method, 34, 46, 83, 111, 131, 283
Motoo's theorem, 264
Moving average, 151, 165
μ-Riemann integrable function, 35
Multiple Wiener integral, 319
Multistep scheme, 279

Narrow convergence, 34, 35, 49
Narrow topology, 35
Negative definite function, 200, 202, 203, 209, 215
Newtonian potential theory, 242
Nietzsche Friedrich, 46, 183
Nonnegative definite, 9, 149
Non-Riemann integrable, 114, 287
Normal number, 124
Novikov's condition, 30
Numerical balayage, 246

One-sided stable process of order $\frac{1}{2}$, 208
One-sided stable subordinator of order α, 218
Optional sampling theorem, 23, 191
ORF, 146
Ornstein–Uhlenbeck process, 28
Ornstein–Uhlenbeck semi-group, 142
Orthogonal projection, 4, 137
Orthogonal random field, 146

Orthogonal stochastic measure, 143, 157
OSM, 143, 148, 151, 154, 157

$\mathbb{P}$-complete, 24
Partial differential equation, 156
Physical quantities (computation of), 227
Pinned Brownian motion, 28
POI, 146, 159, 161
Poincaré's balayage, 241
Poincaré Henri, 241, 247
Point by point calculation, 106
Pointer, 122
Pointwise ergodic theorem, 18, 73, 121, 319
Poisson integral, 242
Poisson kernel, 190
Poisson law, 216
Poisson point process, 224
Poisson process, 204, 205
Polish space, 3, 34, 113, 301
Polya urn scheme, 115
Polynomial approximation, 98
Portfolio, 165
Positive definite, 3
Positive definite symmetric matrix, 100
Potential kernel, 232, 244
Potential theory, 230
Presentation, 301
Pricing by simulation, 265
Prime number, 75, 77
Process with orthogonal increments, 145
Process with stationary independent increments, 161, 198
Product measure, 7
Product σ-field, 7
Progressively measurable, 22
Prokhorov, 35
Prokhorov's lemma, 53
Pseudo-random number, 48, 84
Pseudo-random sequence, 46, 84
PSII, 161, 198, 200, 202, 203, 205, 209, 215
p-uniformly distributed sequence, 131

q.e., 138
Quadratic covariation process, 21
Quadratic risk, 266
Quadratic variation process, 21
Quantum physics, 241
Quasi-everywhere, 138
Quasi-Monte Carlo method, 34, 46, 84, 294
Quasi-Poisson field, 226
Quotient topology, 4

Radial process, 287
Radon measure, 9, 200
Random process, 14
Random variable, 11
Rate of convergence, 46, 58
Recursive simulation, 183, 218
Reduced function, 236, 250
Reduced function algorithm, 250
Réduite, 251, 253, 256
Réduite algorithm, 252, 260
Réduite property, 252, 257
Reflection principle, 190
Regular version of the conditional distribution, 98
Rejection method, 47, 83, 216, 298
Relatively compact, 4
Representation, 235
Residue, 265
Resolvent equation, 232
Resolvent family, 231, 236, 239, 262
Resolvent family (generalized), 237
Restriction of μ to A, 7
Riemann integrable function, 35, 67, 73, 79, 82, 108, 298
Riesz representation theorem, 9
Right process, 262
Robbins–Monro procedure, 58

Rohlin–Halmos lemma, 18, 125

Scaling, 26, 320
Schauder basis, 327
Schwartz distribution, 10
Semi-group-supermedian function, 255
Semi-martingale, 24, 264, 308
Set of potential zero, 258
s-functional, 301, 303
Shift method, 124, 125, 130, 283, 319
Shift operator, 17, 122, 133, 233
Shift transformation, 33
Siegel, 48
σ-finite measure, 7
σ-finite OSM, 145, 156
Simple random variable, 297, 298, 299, 303
Simulatable functional, 296
Simulation, 38, 82, 100, 113, 114, 211, 228
Spectral characteristic, 150
Spectral density, 150, 159, 167, 174
Spectral measure, 150, 151, 156, 157, 158, 167, 173
Spectral representation, 147, 150, 153, 178
Square integrable martingale, 21, 263
Stable law, 209
Stable process, 208, 216
Stable PSII, 205, 209
Standard Brownian motion, 26
Stationary germ, 173
Stationary field, 142
Stationary process, 142
Stochastic approximation, 58
Stochastic differential equation, 269, 274, 287, 295, 306, 319
Stochastic integral, 24, 144, 273, 287, 309
Stone–Weierstrass theorem, 5, 45, 53, 74
Stopped process, 23
Stopping criteria, 54, 64
Stopping time, 20, 22, 122, 234, 235
Stratonovich stochastic integral, 25, 306, 308
Strictly stationary field, 160
Strictly stationary process, 180
Strong invariance principle, 53
Strong law of large numbers, 13, 88, 92
Strong Markov property, 234, 235
Strongly continuous semi-group, 214, 231
Sub-Markov kernel, 16, 230, 244, 250
Submartingale, 19, 21
Subordinator, 161, 188, 198, 202, 204, 205, 210, 213
Subordination in the Bochner sense, 161, 213, 219
Superharmonic function, 242
Supermartingale, 19, 21, 61, 235, 254
Supermedian function, 255
Sweeping, 241
Symbolic expansion, 306
Symmetric difference, 2

Test function, 10, 41
Tight, 35
Topological space, 3
Torus, 4, 72, 81
Total, 4, 99
Transfer function, 150
Transition-density function of the Brownian motion, 27
Transition function, 16
Translation operator, 245

u.d., 44, 81, 83, 114, 118
Uniformly distributed sequence, 44, 131

Universally measurable function, 235, 264
Upper semi-continuous, 36
usc, 36

Van der Corput criterion, 131
Van der Corput sequence, 83
Variation, 67
Vehicle, 110

Weierstrass theorem, 5, 99
Weyl criterion, 71, 132
Weyl Hermann, 45
Weyl polynomial, 132
White noise process, 31, 156, 169, 279
Wiener chaos, 180, 307, 312, 325
Wiener integral, 143
Wiener iterated integral, 278
Wiener multiple integral, 293, 319, 323
Wiener measure, 32, 180
Wiener Norbert, 38, 48, 180
Wiener–Siegel polychotomy algorithm, 48
Wiener space, 32, 269, 284, 287, 293, 294, 301, 302, 320, 326

X-admissible, 150
$x \wedge y$, 1
$x \vee y$, 1
x^+, 1
x^-, 1
$|x|$, 2
$[x]$, 2
$\{x\}$, 2
$]]x, y]]$, 2
$[[x, y]]$, 2

Young modulus, 227

Applied Probability and Statistics (Continued)

BUNKE and BUNKE • Nonlinear Regression, Functional Relations and Robust Methods: Statistical Methods of Model Building
CHATTERJEE and HADI • Sensitivity Analysis in Linear Regression
CHATTERJEE and PRICE • Regression Analysis by Example, *Second Edition*
CLARKE and DISNEY • Probability and Random Processes: A First Course with Applications, *Second Edition*
COCHRAN • Sampling Techniques, *Third Edition*
*COCHRAN and COX • Experimental Designs, *Second Edition*
CONOVER • Practical Nonparametric Statistics, *Second Edition*
CONOVER and IMAN • Introduction to Modern Business Statistics
CORNELL • Experiments with Mixtures, Designs, Models, and the Analysis of Mixture Data, *Second Edition*
COX • A Handbook of Introductory Statistical Methods
*COX • Planning of Experiments
CRESSIE • Statistics for Spatial Data, *Revised Edition*
DANIEL • Applications of Statistics to Industrial Experimentation
DANIEL • Biostatistics: A Foundation for Analysis in the Health Sciences, *Fifth Edition*
DAVID • Order Statistics, *Second Edition*
DEGROOT, FIENBERG, and KADANE • Statistics and the Law
*DEMING • Sample Design in Business Research
DILLON and GOLDSTEIN • Multivariate Analysis: Methods and Applications
DODGE and ROMIG • Sampling Inspection Tables, *Second Edition*
DOWDY and WEARDEN • Statistics for Research, *Second Edition*
DRAPER and SMITH • Applied Regression Analysis, *Second Edition*
DUNN • Basic Statistics: A Primer for the Biomedical Sciences, *Second Edition*
DUNN and CLARK • Applied Statistics: Analysis of Variance and Regression, *Second Edition*
ELANDT-JOHNSON and JOHNSON • Survival Models and Data Analysis
EVANS, PEACOCK, and HASTINGS • Statistical Distributions, *Second Edition*
FISHER and VAN BELLE • Biostatistics: A Methodology for the Health Sciences
FLEISS • The Design and Analysis of Clinical Experiments
FLEISS • Statistical Methods for Rates and Proportions, *Second Edition*
FLEMING and HARRINGTON • Counting Processes and Survival Analysis
FLURY • Common Principal Components and Related Multivariate Models
GALLANT • Nonlinear Statistical Models
GROSS and HARRIS • Fundamentals of Queueing Theory, *Second Edition*
GROVES • Survey Errors and Survey Costs
GROVES, BIEMER, LYBERG, MASSEY, NICHOLLS, and WAKSBERG • Telephone Survey Methodology
HAHN and MEEKER • Statistical Intervals: A Guide for Practitioners
HAND • Discrimination and Classification
*HANSEN, HURWITZ, and MADOW • Sample Survey Methods and Theory, Volume I: Methods and Applications
*HANSEN, HURWITZ, and MADOW • Sample Survey Methods and Theory, Volume II: Theory
HEIBERGER • Computation for the Analysis of Designed Experiments
HELLER • MACSYMA for Statisticians
HINKELMAN and KEMPTHORNE: Design and Analysis of Experiments, Volume I: Introduction to Experimental Design
HOAGLIN, MOSTELLER, and TUKEY • Exploratory Approach to Analysis of Variance
HOAGLIN, MOSTELLER, and TUKEY • Exploring Data Tables, Trends and Shapes
HOAGLIN, MOSTELLER, and TUKEY • Understanding Robust and Exploratory Data Analysis

*Now available in a lower priced paperback edition in the Wiley Classics Library.

Applied Probability and Statistics (Continued)

HOCHBERG and TAMHANE • Multiple Comparison Procedures
HOEL • Elementary Statistics, *Fifth Edition*
HOGG and KLUGMAN • Loss Distributions
HOLLANDER and WOLFE • Nonparametric Statistical Methods
HOSMER and LEMESHOW • Applied Logistic Regression
IMAN and CONOVER • Modern Business Statistics
JACKSON • A User's Guide to Principle Components
JOHN • Statistical Methods in Engineering and Quality Assurance
JOHNSON • Multivariate Statistical Simulation
JOHNSON and KOTZ • Distributions in Statistics
Discrete Distributions
Continuous Univariate Distributions—1
Continuous Univariate Distributions—2
Continuous Multivariate Distributions
JUDGE, GRIFFITHS, HILL, LÜTKEPOHL, and LEE • The Theory and Practice of Econometrics, *Second Edition*
JUDGE, HILL, GRIFFITHS, LÜTKEPOHL, and LEE • Introduction to the Theory and Practice of Econometrics, *Second Edition*
KALBFLEISCH and PRENTICE • The Statistical Analysis of Failure Time Data
KASPRZYK, DUNCAN, KALTON, and SINGH • Panel Surveys
KISH • Statistical Design for Research
KISH • Survey Sampling
LAWLESS • Statistical Models and Methods for Lifetime Data
LEBART, MORINEAU, and WARWICK • Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices
LEE • Statistical Methods for Survival Data Analysis, *Second Edition*
LEPAGE and BILLARD • Exploring the Limits of Bootstrap
LEVY and LEMESHOW • Sampling of Populations: Methods and Applications
LINHART and ZUCCHINI • Model Selection
LITTLE and RUBIN • Statistical Analysis with Missing Data
MAGNUS and NEUDECKER • Matrix Differential Calculus with Applications in Statistics and Econometrics
MAINDONALD • Statistical Computation
MALLOWS • Design, Data, and Analysis by Some Friends of Cuthbert Daniel
MANN, SCHAFER, and SINGPURWALLA • Methods for Statistical Analysis of Reliability and Life Data
MASON, GUNST, and HESS • Statistical Design and Analysis of Experiments with Applications to Engineering and Science
McLACHLAN • Discriminant Analysis and Statistical Pattern Recognition
MILLER • Survival Analysis
MONTGOMERY and PECK • Introduction to Linear Regression Analysis, *Second Edition*
NELSON • Accelerated Testing, Statistical Models, Test Plans, and Data Analyses
NELSON • Applied Life Data Analysis
OCHI • Applied Probability and Stochastic Processes in Engineering and Physical Sciences
OKABE, BOOTS, and SUGIHARA • Spatial Tesselations: Concepts and Applications of Voronoi Diagrams
OSBORNE • Finite Algorithms in Optimization and Data Analysis
PANKRATZ • Forecasting with Dynamic Regression Models
PANKRATZ • Forecasting with Univariate Box-Jenkins Models: Concepts and Cases
PORT • Theoretical Probability for Applications
PUTERMAN • Markov Decision Processes: Discrete Stochastic Dynamic Programming
RACHEV • Probability Metrics and the Stability of Stochastic Models
RÉNYI • A Diary on Information Theory

*Now available in a lower priced paperback edition in the Wiley Classics Library.

Applied Probability and Statistics (Continued)

RIPLEY • Spatial Statistics
RIPLEY • Stochastic Simulation
ROSS • Introduction to Probability and Statistics for Engineers and Scientists
ROUSSEEUW and LEROY • Robust Regression and Outlier Detection
RUBIN • Multiple Imputation for Nonresponse in Surveys
RYAN • Statistical Methods for Quality Improvement
SCHUSS • Theory and Applications of Stochastic Differential Equations
SCOTT • Multivariate Density Estimation: Theory, Practice, and Visualization
SEARLE • Linear Models
SEARLE • Linear Models for Unbalanced Data
SEARLE • Matrix Algebra Useful for Statistics
SEARLE, CASELLA, and McCULLOCH • Variance Components
SKINNER, HOLT, and SMITH • Analysis of Complex Surveys
STOYAN • Comparison Methods for Queues and Other Stochastic Models
STOYAN, KENDALL, and MECKE • Stochastic Geometry and Its Applications
THOMPSON • Empirical Model Building
THOMPSON • Sampling
TIERNEY • LISP-STAT: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics
TIJMS • Stochastic Modeling and Analysis: A Computational Approach
TITTERINGTON, SMITH, and MAKOV • Statistical Analysis of Finite Mixture Distributions
UPTON and FINGLETON • Spatial Data Analysis by Example, Volume I: Point Pattern and Quantitative Data
UPTON and FINGLETON • Spatial Data Analysis by Example, Volume II: Categorical and Directional Data
VAN RIJCKEVORSEL and DE LEEUW • Component and Correspondence Analysis
WEISBERG • Applied Linear Regression, *Second Edition*
WESTFALL and YOUNG • Resampling-Based Multiple Testing: Examples and Methods for *p*-Value Adjustment
WHITTLE • Optimization Over Time: Dynamic Programming and Stochastic Control, Volume I and Volume II
WHITTLE • Systems in Stochastic Equilibrium
WONNACOTT and WONNACOTT • Econometrics, *Second Edition*
WONNACOTT and WONNACOTT • Introductory Statistics, *Fifth Edition*
WONNACOTT and WONNACOTT • Introductory Statistics for Business and Economics, *Fourth Edition*
WOODING • Planning Pharmaceutical Clinical Trials: Basic Statistical Principles
WOOLSON • Statistical Methods for the Analysis of Biomedical Data

Tracts on Probability and Statistics

ergence of Probability Measures
r Information in Linear Models

ISBN 0-471-54641-0